Journal of Applied Psychology

Edited by

Donald G. Paterson
University of Minnesota

Consulting Editors

Paul S. Achilles, *Psychological Corporation;* Walter V. Bingham, *A.G.O., War Department;* Harold E. Burtt, *Ohio State University;* Arthur I. Gates, *T. C. Columbia University;* John G. Jenkins, *University of Maryland;* Irving Lorge, *T. C. Columbia University;* Quinn McNemar, *Stanford University;* Willard C. Olson, *University of Michigan;* James P. Porter, *Swarthmore, Pennsylvania;* Edward K. Strong, Jr., *Stanford University;* Morris S. Viteles, *University of Pennsylvania;* Joseph Zubin, *N. Y. Psychiatric Institute.*

Volume 29, 1945

Published Bi-monthly by The American Psychological Association, Inc.
With the Cooperation of The American Association for Applied Psychology
Prince and Lemon Sts., Lancaster, Pa., and Northwestern University, Evanston, Illinois

Copyright, 1945, by The American Psychological Association, Inc.

Reprinted with the permission of the original publishers.

Contents of Volume 29

Articles

Barkley, K. L. Values Students Reported from the Study of Emotions ... 378
Barrett, D. M. Aptitude and Interest Patterns of Art Majors in a Liberal Arts College ... 483
Berdie, R. F. Range of Interests ... 268
Capwell, D. F. Personality Patterns of Adolescent Girls: I. Girls Who Show Improvement in IQ ... 212
Capwell, D. F. Personality Patterns of Adolescent Girls: II. Delinquents and Non-Delinquents ... 289
Corsini, R. A New Method for the Administration of Individual Intelligence Tests ... 356
Eisenberg, P. Two Methods of Combining Attitudes of Like, Indifference and Dislike into One Score ... 246
England, A. O. Military Job Evaluation: Army Air Forces ... 437
File, Q. W. The Measurement of Supervisory Quality in Industry ... 323
Forlano, G., and Kirkpatrick, F. H. Intelligence and Adjustment Measurements in the Selection of Radio Tube Mounters ... 257
Griffiths, G. R. The Relationship Between Scholastic Achievement and Personality Adjustment of Men College Students ... 360
Guest, L. Magazine vs. Personal Interview Votes in the Consumer Jury Advertising Test ... 399
Hahn, M. E., and Williams, C. T. The Measured Interests of Marine Corps Women Reservists ... 198
Harmon, L. R., and Wiener, D. N. Use of the Minnesota Multiphasic Personality Inventory in Vocational Advisement ... 132
Henderson, M. T., Crews, A., and Barlow, J. A Study of the Effect of Music Distraction on Reading Efficiency ... 313
Henning, E. J., and Carter, H. D. Participation in High-School Football as a Factor Affecting College Attendance and Scholarship ... 236
Hildreth, G. A School Survey of Eye-Hand Dominance ... 83
Hildreth, H. M. Single-Item Tests for Psychometric Screening ... 262
Horn, C. A., and Smith, L. F. The Horn Art Aptitude Inventory ... 350
Hunt, H. F. A Note on the Problem of Brain Damage in Rehabilitation and Personnel Work ... 282
Jacobson, W. E. First Impressions of Classmates ... 142

Keller, F. S. Studies in International Morse Code. 4. A Note on Second-Level Training in Code Reception.................... 161
Knower, F. H. Studies in the Symbolism of Voice and Action: V. The Use of Behavioral and Tonal Symbols as Tests of Speaking Achievement.. 229
Lawshe, C. H., Jr. Studies in Job Evaluation: II. The Adequacy of Abbreviated Point Ratings for Hourly-Paid Jobs in Three Industrial Plants.. 177
Lawshe, C. H., Jr., and Tiffin, J. The Accuracy of Precision Instrument Measurement.. 413
Lennon, R. T., and Baxter, B. Predictable Aspects of Clerical Work 1
Lindahl, L. G. Movement Analysis as an Industrial Training Method.. 420
Link, H. C. The Psychological Corporation Index of Public Opinion 103
Long, L. Relationship Between Interests and Abilities: A Study of the Strong Vocational Interest Blank and the Zyve Scientific Aptitude Test.. 191
Luborsky, L. Aircraft Recognition: I. The Relative Efficiency of Teaching Procedures... 385
Luborsky, L. Aircraft Recognition: II. A Study of Prognostic Tests 449
Macdonald, G. L. Measuring Progress in Radio Training........ 301
Martin, A. H. A Worry Inventory.............................. 68
Martin, H. G. The Construction of the Guilford-Martin Inventory of Factors G-A-M-I-N...................................... 208
McPherson, M. W. A Method of Objectively Measuring Shop Performance.. 22
McQuitty, L. L. Psychological Principles in Army Administration.. 458
Philip, B. R. A Method for Investigating Color Preferences in Fashions.. 108
Portenier, L. G. Mechanical Aptitudes of University Women..... 477
Schenet, N. G. An Analysis of Absenteeism in One War Plant... 27
Schmidt, H. O. Test Profiles as a Diagnostic Aid: The Minnesota Multiphasic Inventory.................................... 115
Shuman, J. T. The Value of Aptitude Tests for Factory Workers in the Aircraft Engine and Propeller Industries................. 150
Shuman, J. T. The Value of Aptitude Tests for Supervisory Workers in the Aircraft Engine and Propeller Industries............ 185
Stalnaker, J. M. Personnel Placement in the Armed Forces...... 338
Steel, M., Balinsky, B., and Lang, H. A Study on the Use of a Work Sample.. 14
Stump, N. F. A Statistical Study of Visual Functions and Industrial Safety.. 467

Sumner, F. C., and Shaed, D. L. **Negro-White Attitudes Towards the Administration of Justice as Affecting Negroes** 368
Tinker, M. A. **Effect of Visual Adaptation Upon Intensity of Illumination Preferred for Reading With Direct Lighting** 471
Toolan, W. T. **Merit Examination Cut-Offs and Weights** 493
Tyler, L. E. **Relationships Between Strong Vocational Interest Scores and Other Attitude and Personality Factors** 58
Wilson, G. M., and Staff. **Adapting the Minnesota Rate of Manipulation Test to Factory Use** 346
Wulfeck, W. H. **The Role of the Psychologist in Market and Advertising Research** 95
Zeligs, R. **Social Factors Annoying to Children** 75
Zubin, J., and Peatman, J. G. **Testing the Pulling Power of Advertisements by the Split-Run Copy Method** 40

Book Reviews

Bowley's Guiding the Normal Child: Glenn M. Blair 504
Cason's Mechanical Methods for Increasing the Speed of Reading: Leona E. Tyler 319
Cole's Attaining Maturity: Miles A. Tinker 90
Dvorine's Color Perception Testing Charts (Vol. 1); Color Perception Training Charts (Vol. 2): Miles A. Tinker 175
Gallup's A Guide to Public Opinion Polls: Donald G. Paterson 174
Hahn's Stuttering, Significant Theories and Therapies: Franklin H. Knower 89
Halsey's Making and Using Industrial Service Ratings: Charles C. Gibbons 318
Kingsley's Recruiting Applicants for the Public Service. A Report Submitted to the Civil Service Assembly by the Committee on Recruiting Applicants for the Public Service: Arthur Burton . . . 90
Klein's Mental Hygiene, The Psychology of Personal Adjustment: Fred McKinney 502
Luckiesh's Light, Vision and Seeing: Miles A. Tinker 252
MacKintosh's The War and Mental Health in England: John E. Anderson 410
McMurry's Handling Personality Adjustment in Industry: Charles C. Gibbons 173
Melville's Color Vision: F. L. Dimmick 409
Rosenstein's The Scientific Selection of Salesmen: Milton E. Hahn . 498
Sargent's The Basic Teachings of the Great Psychologists: Harold E. Burtt 320

Scott, Morgan, and Lehman's Developing a Student Guidance Program in an Instructional Department: Clara M. Brown........ 501
Traxler's Techniques of Guidance: Tests, Records, and Counseling in a Guidance Program: F. S. Beers.......................... 499
Woolpert's Municipal Personnel Administration: John K. McKay.. 497

Miscellaneous

Job Specifications in Applied Psychology........................ 164
New Books, Monographs, and Pamphlets.......176, 254, 321, 411, 506

Journal of Applied Psychology

Vol. 29, No. 1 February, 1945

Predictable Aspects of Clerical Work *

Roger T. Lennon

Research and Test Service, World Book Company

and

Brent Baxter

Ohio State University

The usual procedure in evaluating the efficiency of aptitude tests in the prediction of clerical work is to correlate the test scores with some measure or rating of success in clerical work. Few, if any, attempts, however, have been made to determine what specific aspects of clerical work are related to scores on the types of tests most commonly used in selecting clerical workers. This article describes one such attempt, carried out in connection with the usual-type investigation of the validity of two tests used for predicting clerical success. The study described was conducted in a large government agency employing several thousand clerical workers. The plan was to have a group of clerical employees rated on a series of specific statements concerning different aspects of clerical work, and to note on which statements ratings were related significantly to test scores.

The Tests. The tests involved in this study were as follows:

a. Learning Ability Test. This is a local adaptation of Army Alpha, Kansas Revision (by Schrammel and Wood); it includes six subtests: arithmetic problems, common sense, same-opposites, rearranged sentences, analogies, and general information, requiring 27 minutes working time in all. It correlates .88 with the Otis Quick-Scoring Test of Mental Ability-Gamma Test. "Learning Ability Test" is the designation used locally, and hereinafter, to refer to this test.

b. Clerical Aptitude Test. This is a test developed by the authors to measure speed and accuracy in the performance of simple clerical tasks.

---

* Acknowledgment is gratefully made to Miss Evelyn Potechin for assistance in the statistical computations.

It is also composed of six subtests: alphabetizing, number checking, coding, digit counting, arithmetic computation, and locating information from tabular data. The test requires 21 minutes of working time. Correlations of .67 with the Civil Service Commission Clerical Examination, and .65 with the Learning Ability Test have been obtained.

Both tests were administered to practically all applicants for clerical positions in the agency at the time of employment, but virtually no applicants were rejected on the basis of the test results.

The Subjects. The employees chosen for the study included all who had the following qualifications: (1) had taken both tests at the time of employment; (2) had been on the job for at least three months; and (3) were doing fairly simple clerical work, involving no supervisory responsibilities. The group originally chosen for study numbered about 250, and comprised about 50% clerks of various kinds, 30% stenographers, 10% typists, and 10% miscellaneous workers such as clock card checkers and bookkeeping machine operators.[1]

The Check List. One of the criteria which was used to evaluate the tests as predictors of clerical success was a so-called Clerical Check List. This Check List consisted of a series of statements concerning the work of clerical employees. The method used in developing the Check List followed in many ways a method originally proposed by Probst.[2] The validity of this method as a technique for measuring success in clerical work is not discussed herein, since this article deals only with the individual items in the List, and not with any composite rating. It is not assumed that the statements in the List are independent. The entire Check List appears below, together with the directions for filling it out which accompanied it. It will be noted that the List includes both favorable and unfavorable comments on the quality and quantity of the employee's work. The items are both specific and general; that is, some statements are pertinent to only a specific task, such as the transcribing of dictation notes, while other statements pertain to characteristics such as neatness, orderliness, etc., which apply to all kinds of work.

Check List for Clerical Employees

Directions: Below is a series of statements which have been made about clerical employees. Read each statement carefully, and decide whether or not it is a true statement about the employee under consideration. Then put a mark in the margin at the left of the statement in accordance with the directions below.

a. If you believe the statement is true, put a plus sign (+) before the statement.

---

[1] For the information of those familiar with the classification system in Federal agencies, these subjects may be identified as grade CAF-1, 2, or 3.

[2] J. B. Probst, *Personnel*, 1931, 8, 20-24.

b. If you believe the statement is not true, put a "0" before it.

c. If you do not have sufficient knowledge about the employee to decide about the statement, mark it with a question mark (?).

d. If the statement is not pertinent to the work the employee is doing, mark it with "NP." For example, if the statement says something about typing, and the employee being considered does no typing, mark the statement with "NP."

e. Place one of the above marks at the left of each statement. There may be some special event which has helped you to decide about a statement or some special information. In these cases do not hesitate to note it in the column under "Comments." Code for Marking Statements: True = +; Not true = 0; Don't know = ?; and Not pertinent = NP.

Mark here — Statements

.... 1. Uses the correct printed forms for each operation in his work.
.... 2. Checks on his work have revealed that he makes very few errors.
.... 3. Can arrange tabulations and statistical memoranda in an easily understood form.
.... 4. Attends systematically to matters that require periodic attention.
.... 5. Keeps his records up to date.
.... 6. Is alert to notice the improper routing of a file or letter.
.... 7. Gets important letters out before the end of a day instead of holding them over night.
.... 8. Checks his work for errors before releasing it.
.... 9. Accurately transfers information to forms.
............10. Has overlooked obvious errors in the work he is handling.
............11. Does a fair share of the work in his unit.
............12. Often does necessary but unrequested work on his own initiative.
............13. Transmits messages correctly and intelligibly.
............14. Has had to be corrected for the same error more than once.
............15. Has sufficient knowledge of the work he is handling to detect obvious errors.
............16. Rarely misspells a word.
............17. Some of the stencils cut have had to be retyped.
............18. Sometimes forgets matters which should receive prompt attention.
............19. Sometimes gets lost in routine and forgets important details.
............20. Knows where to go for necessary information.
............21. Is accurate in computations and calculations.
............22. Takes the easiest way when it means the results will be incomplete and not wholly satisfactory.
............23. Work has sometimes had to be assigned to others because he is slow.
............24. Can, when required, compose a letter or memorandum that covers all pertinent points and presents all necessary details.
............25. A "peak" in work to be done discourages him to the detriment of his work.
............26. Resumes work promptly after an interruption.
............27. Turns in letters which reveal errors in English which were overlooked.
............28. Tries to avoid added responsibilities.
............29. Does less work than other people in his unit doing the same kind of work.
............30. Correctly compiles material.
............31. Typed letters are occasionally marred by erasures or strikeovers.

Mark here	Statements
............32.	Is eager to learn more about the work of his unit.
............33.	His working instructions have to be repeated frequently.
............34.	Avoids unnecessary duplication in his work.
............35.	Sometimes fails to have adequate supplies and materials on hand.
............36.	Has had to do some of his work over because he has started before clearly understanding what is to be done.
............37.	Carbon copies have sometimes shown folds, creases, smudges, etc.
............38.	If he is assigned more than one job at a time, he becomes confused.
............39.	He is familiar with the relation of his work to other work in the branch.
............40.	Has started some new assignments without a clear understanding of what is to be done.
............41.	On his own initiative he has obtained outside training which has improved his effectiveness as a worker.
............42.	Calls attention to possible errors in materials before transcribing them.
............43.	Does a fair share of the more difficult tasks within the unit.
............44.	Checks on his work reveal that he makes no mistakes.
............45.	The speed of note-taking (dictation) is as fast as the work demands.
............46.	Has used poor grammar in typing letters from general directions.
............47.	When he has to speed up because of a peak load, there is an increase in his percentage of errors.
............48.	Has made helpful suggestions about work handled.
............49.	Some work has had to be returned to him to be done over because of the quality of the work.
............50.	Copies of work show an uneven typing touch.
............51.	Reviews his work for discrepancies before turning it in.
............52.	Correctly uses the terminology of his job.
............53.	Difficult tasks challenge rather than confuse him.
............54.	Has sometimes made the wrong number of copies.
............55.	Has misdirected mail when sorting it for office distribution.
............56.	Is often entrusted with tabulation of difficult materials.
............57.	Quickly learns where and how to locate people whom the supervisor often calls.
............58.	Returns all letters, records, files, etc., to their proper place promptly.
............59.	Distracts or interrupts a dictator unnecessarily.
............60.	Is confused by sudden changes in assignment.
............61.	Bothers others by asking them how to spell words.
............62.	Typed work is always orderly and well aligned.
............63.	Does not turn out as much work as the other members of the unit.
............64.	Grasps new rules, regulations and procedures quickly.
............65.	Is unable to carry on effectively more than one task at a time.
............66.	Is an employee to whom the most difficult assignments can be given.
............67.	Works equally well with or without close supervision.
............68.	Dictators have to slow up because of his inability to take dictation fast enough.
............69.	Produces work of acceptable quality even under pressure.
............70.	Accurately ascertains the business of visitors and directs them correctly.
............71.	Adapts himself easily and quickly when given a new assignment.
............72.	Cannot take full stenographic notes so that other stenographers can transcribe from them if necessary.

Mark here	Statements
............73.	Checks and corrects such items as names, titles, addresses, file references, dates, etc., before releasing correspondence.
............74.	Desk and files are neat and orderly.
............75.	Can analyze records quickly so as to locate needed information.
............76.	Is frequently asked to make up special memoranda.
............77.	Does not always set up letters in accordance with regulations.
............78.	Is able to maintain the average rate of production of the unit.
............79.	Does not detect inconsistencies in the material typed or transcribed.
............80.	Reminds the supervisor of his appointments if he seems to have overlooked them.
............81.	Does not remember frequently-used names and numbers.
............82.	Because of his slowness you sometimes hesitate to assign some jobs to him.
............83.	Has demonstrated ability to compose acceptable letters or memoranda.
............84.	In transcribing dictaphone rolls, can make reasonable interpretations where the roll is not clear.
............85.	Passes on blame for own errors.
............86.	Is able to concentrate on duties despite the frequent interruptions which come with acting as receptionist.
............87.	Has been known to make unnecessary duplication in his work.
............88.	His working instructions have to be repeated.
............89.	In the use of carbons with forms, the carboned material appears in the proper spaces.
............90.	Is inclined to sacrifice accuracy for speed.

Copies of the Check List, with directions as shown below, were sent to supervisors of the group selected for follow-up.

Directions for Supervisors Marking the Check Lists

1. The Personnel Section is making a study of the effectiveness with which its aptitude tests predict the efficiency of its clerical personnel. In order to complete this study the Section is requesting that supervisors mark Check Lists like the one attached for some of their employees. Only those clerical employees who have been tested and who have been on the job for at least three months are being studied. These records will not become part of the permanent files nor will they affect the Civil Service efficiency ratings.

2. It is urged that careful attention be given to checking these lists for otherwise they are worthless. Please read and follow carefully the directions on the first page of the Check List. It is suggested that all of the statements on the List be read before any of them are marked. About fifteen minutes is required to complete a Check List.

3. The name of the employee to be checked is in the upper right hand corner of the List. The name of the supervisor checking the List should be written in the upper left-hand corner of the first page. The number of weeks during which the work of the employee has been reviewed by the supervisor indicated should be written beneath the name of the supervisor.

4. It is believed that there is no need to discuss the Check List with the employee unless it is specially desired. In any case it should not be discussed until after the checking has been done.

5. Please return these Check Lists completed to your Branch Chief within four days.

No supervisor was asked to complete Check Lists for more than four employees and most supervisors had to mark only one. Every effort was made to have the checking done by the immediate supervisor. Most of the reporting supervisors had about fifteen employees under their direction. As the Check Lists were returned, each one was reviewed to see that the marking had been done according to directions, and that the proper supervisor had marked it. Of the Check Lists sent out, 80% were returned. The chief reasons for failure of all the Check Lists to be returned were separations of the employees or the supervisors, and the transferring of employees to different supervisors.

It is realized that highly accurate ratings were not in all cases obtained from this group of supervisors, many of whom were untrained in checking the performance of employees. The supervisors' checking was internally consistent, however, the correlation being estimated as at least .85. Some evidence was obtained on the agreement between supervisors on the total Check List score. A group of 62 employees were rated by two supervisors, the correlation between pairs of supervisors' ratings being .69. No evidence is available on the reliability or validity of the marking of the individual items. Since these Check List ratings leave much to be desired as a criterion, any findings as to the number of items predicted significantly by the tests are likely to err on the conservative side.

Results

After the Check Lists had been filled out by the supervisors in accordance with the procedure described above, an analysis was undertaken to determine which of the items were significantly related to scores on the tests given at the time of employment; that is, which ones could have been predicted, to some extent at least, on the basis of test scores. From the total group for which Check Lists were completed, a group with high scores and a second group with low scores on the Learning Ability Test, constituting the highest and lowest 27%, were selected. The number of cases in each group was 58. A tabulation of responses to each item was made for these two groups, yielding for each item the number in the group for which a "+," "0," "NP," or "?" had been recorded. The results of the tabulation appear in Table 1.

The statistical analysis in this paper deals only with the proportion of the high and low groups for which each item was marked "+," i.e., true. The proportion considered was the relation of the number in the group having the item marked "+" to the sum of the "+" and "0" markings. Thus allowance was made for the "?" responses, reflecting inadequate knowledge on the part of the supervisors, and for the "NP" responses, so that the proportion for any item is based only on those individuals for whom the item is appropriate—e.g., typing items for typists. It was sometimes found that the proportion of "NP" 's marked for an item was significantly different for high and low scoring groups, reflecting differences in the kind of work assigned to low and high scoring people. Inasmuch as this study is interested in the prediction of quality and quantity of work rather than the kind of work, these differences will not be discussed here.

The chi-square test was used to evaluate the significance of the differences between the high and low groups, and those differences for which P was less than .10 were regarded as significant—in other words, these differences which were at the 10 per cent level of significance.[3] In the calculation of χ^2, Yates' correction for continuity[4] was uniformly used to provide a more refined test even when the numbers of cases were sufficiently large so that it was not strictly necessary. Use of this technique is illustrated by the following example, which is the calculation of chi-square for Item 24 for the high and low groups on the Learning Ability Test.

	High Score	Low Score	Total
+	20	11	31
0	4	6	10
Total	24	17	41

$$\chi^2 = \frac{\left(20 \times 6 - 11 \times 4 - \frac{41}{2}\right)^2 41}{24 \times 17 \times 10 \times 31} = .998$$

For $\chi^2 = .998$, with one degree of freedom, P equals approximately .30. Therefore, according to our standard, this difference is not significant.

On the basis of this analysis, 12 items were found to differentiate significantly between the high and the low groups on the Learning Ability Test. These items are indicated in Table 1.

The same procedure was followed in determining which items were significantly related to scores on the Clerical Aptitude Test. High-scoring and low-scoring groups on this test, again constituting the highest and lowest 27% of the total group, were selected, and tabulations of the responses made for these groups. Because there is a substantial corre-

---

[3] In the actual computation of χ^2, it is not necessary to compute the proportions as such, since the calculations are performed in terms of frequencies; hence these proportions are not included in Table 1.

[4] See Goulden, C. H., *Methods of statistical analysis.* New York: John Wiley and Sons, 1939, pp. 102–106.

Table 1

Summary of Responses to Check-List Items for High-Scoring and Low-Scoring Groups on Learning Ability and Clerical Aptitude Tests

	Learning Ability Test									Clerical Aptitude Test								
Group	Upper 27%				Lower 27%					Upper 27%				Lower 27%				
Marking	+	0	NP	?	+	0	NP	?	Sig.*	+	0	NP	?	+	0	NP	?	Sig.*
Item No.																		
1	50	1	7	0	40	0	18	0	N	55	0	3	0	42	0	16	0	N
2	52	6	0	0	48	9	1	0	N	55	2	0	1	49	8	1	0	N
3	20	0	32	6	18	1	36	3	N	27	1	28	2	18	1	36	3	N
4	41	2	12	4	31	3	24	0	N	43	0	12	3	32	4	22	0	S
5	42	3	12	1	36	0	22	0	N	50	0	7	1	35	0	22	1	N
6	36	5	16	1	24	8	23	3	N	38	14	5	1	21	9	23	5	S
7	32	4	22	0	18	0	40	0	N	36	2	20	0	18	0	40	0	N
8	51	5	1	1	45	4	7	2	N	52	4	2	0	48	5	3	2	N
9	46	1	11	0	33	2	22	1	N	49	1	8	0	36	1	19	2	N
10	11	46	0	1	4	50	3	1	S	6	49	2	1	9	47	2	0	N
11	54	2	1	1	55	2	1	0	N	57	0	0	1	56	2	0	0	N
12	37	15	3	3	26	22	7	3	S	39	13	4	2	25	19	8	6	S
13	38	1	17	2	29	2	27	0	N	39	1	16	2	26	2	30	0	N
14	8	49	0	1	15	39	1	3	S	6	50	0	2	15	38	1	4	S
15	54	4	0	0	48	8	1	1	N	54	3	0	1	49	8	1	0	N
16	44	3	8	3	30	2	25	1	N	42	5	6	5	35	2	20	1	N
17	1	6	49	2	0	5	51	2	N	1	10	44	3	1	6	51	0	N
18	5	47	6	0	7	42	7	2	N	7	47	4	0	7	40	8	3	N
19	8	45	4	1	7	41	8	2	N	8	47	3	0	5	40	11	2	N
20	51	3	1	3	49	6	1	2	N	55	2	0	1	46	7	1	4	S

Table 1—*Continued*

	Learning Ability Test									Clerical Aptitude Test								
Group	Upper 27%				Lower 27%					Upper 27%				Lower 27%				
Marking	+	0	NP	?	+	0	NP	?	Sig.*	+	0	NP	?	+	0	NP	?	Sig.*
Item No.																		
21	15	3	36	4	20	3	33	2	N	23	2	31	2	18	3	36	1	N
22	3	52	1	2	5	47	5	1	N	0	57	0	1	4	46	5	3	S
23	7	51	0	0	8	47	2	1	N	3	55	0	0	9	46	2	1	S
24	20	4	31	3	11	6	34	7	N	22	6	26	4	8	5	38	7	N
25	4	53	0	1	2	51	4	1	N	4	53	0	1	3	50	5	0	N
26	50	7	0	1	51	4	2	1	N	50	7	0	1	48	7	1	2	N
27	9	28	18	3	7	17	32	2	N	7	29	19	3	4	16	36	2	N
28	5	49	3	1	6	49	3	0	N	3	54	1	0	1	51	5	1	N
29	4	50	4	0	8	48	2	0	N	1	57	0	0	7	49	2	0	S
30	39	2	16	1	37	3	17	1	N	39	1	17	1	35	2	17	4	N
31	10	25	22	1	6	12	39	1	N	9	24	24	1	4	12	42	0	N
32	47	6	1	4	47	6	1	4	N	49	6	1	2	45	4	2	7	N
33	7	51	0	0	12	45	0	1	S	5	53	0	0	5	51	0	2	N
34	51	4	2	1	43	9	4	2	N	54	2	1	1	46	8	2	2	S
35	3	45	10	0	3	37	18	0	N	5	46	7	0	3	41	14	0	N
36	5	49	2	2	15	39	4	0	S	5	50	1	2	17	38	1	2	S
37	6	31	21	0	1	16	40	1	N	6	28	23	1	2	13	43	0	N
38	4	44	7	3	11	36	10	1	S	7	45	6	0	11	36	8	3	N
39	48	7	0	3	43	7	3	5	N	50	7	1	0	41	6	5	6	N
40	5	44	8	1	11	44	3	0	N	7	45	5	1	15	37	6	0	S
41	2	19	10	27	7	24	5	22	N	8	21	3	26	9	15	6	28	N
42	39	5	13	1	31	2	25	0	N	44	1	12	1	32	0	26	0	N
43	45	5	7	1	35	11	11	1	S	47	4	6	1	34	11	11	2	S
44	20	33	2	3	14	43	1	0	N	19	35	1	3	17	38	1	2	N
45	27	30	0	1	6	1	50	1	N	24	1	32	1	6	0	52	0	N

Table 1—*Continued*

	Learning Ability Test									Clerical Aptitude Test								
Group	Upper 27%				Lower 27%					Upper 27%				Lower 27%				
Marking	+	0	NP	?	+	0	NP	?	Sig.*	+	0	NP	?	+	0	NP	?	Sig.*
Item No.																		
46	1	27	28	2	2	11	45	0	N	3	26	28	1	0	8	49	1	N
47	11	42	5	0	14	36	6	2	N	10	45	2	1	13	40	3	2	N
48	31	17	6	4	29	21	2	6	N	38	13	4	3	22	23	5	8	S
49	10	42	6	0	16	35	7	0	N	10	46	2	0	14	37	7	0	N
50	4	32	22	0	1	16	41	0	N	1	35	22	0	1	13	44	0	N
51	49	5	4	0	38	6	13	1	N	51	6	1	0	37	5	14	2	N
52	38	3	12	5	31	4	18	5	N	44	2	9	3	31	4	16	7	N
53	36	9	8	5	31	15	7	5	N	42	9	4	3	29	16	8	5	S
54	10	25	22	1	7	15	36	0	N	10	24	22	2	4	15	38	1	N
55	1	13	43	1	1	13	44	0	N	2	15	41	0	0	12	46	0	N
56	13	4	41	0	11	10	37	0	N	20	6	32	0	11	7	40	0	N
57	26	2	29	1	22	6	30	0	N	34	1	22	1	21	5	32	0	S
58	46	4	7	1	40	2	16	0	N	47	4	6	1	39	2	17	0	N
59	4	30	23	1	1	24	33	0	N	4	33	21	0	2	21	35	0	N
60	7	44	7	0	12	36	9	1	N	6	47	5	0	11	35	11	1	N
61	4	45	7	2	1	38	17	2	N	0	47	8	3	2	40	14	2	N
62	32	6	20	0	16	4	38	0	N	35	5	18	0	13	3	42	0	N
63	5	48	4	1	12	40	6	0	S	1	53	2	2	9	46	3	0	S
64	51	4	2	1	44	13	0	1	S	54	3	0	1	40	14	1	3	S
65	5	44	8	1	13	33	10	2	S	7	42	5	4	15	31	10	2	S
66	26	16	11	5	18	30	6	4	S	29	17	8	4	20	29	4	5	S
67	50	8	0	0	46	11	1	0	N	53	5	0	0	47	10	1	0	N
68	2	23	32	1	2	7	49	0	N	2	23	33	0	0	7	51	0	N
69	49	5	3	1	42	7	6	3	N	54	2	1	1	43	6	6	3	N
70	12	1	45	0	9	0	47	2	N	12	1	43	2	10	1	46	1	N

Table 1—*Continued*

Group	Learning Ability Test									Clerical Aptitude Test								
	Upper 27%				Lower 27%					Upper 27%				Lower 27%				
Marking	+	0	NP	?	+	0	NP	?	Sig.*	+	0	NP	?	+	0	NP	?	Sig.*
Item No.																		
71	46	4	6	2	41	7	6	4	N	48	5	2	3	39	9	6	4	N
72	4	17	33	4	3	3	50	2	N	4	19	31	4	0	5	53	0	N
73	35	3	19	1	21	2	35	0	N	36	1	19	2	19	2	36	1	N
74	56	2	0	0	50	2	5	1	N	56	2	0	0	50	3	3	2	N
75	34	4	18	2	31	6	20	1	N	42	3	12	1	31	6	19	2	N
76	7	18	33	0	8	19	31	0	N	23	24	10	1	5	21	32	0	S
77	2	32	23	1	3	15	40	0	N	2	33	22	1	3	12	42	1	N
78	52	3	3	0	46	7	5	0	N	55	1	2	0	47	8	3	0	S
79	8	30	18	2	5	20	33	0	N	3	36	17	2	4	23	31	0	N
80	21	2	34	1	14	4	39	1	N	21	1	35	1	12	3	41	2	N
81	4	45	7	2	8	37	11	2	N	1	46	8	3	7	42	6	3	S
82	8	47	3	0	12	40	6	0	N	6	50	2	0	14	40	4	0	S
83	20	2	32	4	9	4	41	4	N	25	1	28	4	10	3	41	4	N
84	2	0	56	0	0	0	57	1	N	1	0	55	2	1	0	57	0	N
85	5	51	1	1	1	56	1	0	N	3	54	0	1	0	56	1	1	N
86	15	1	42	0	12	2	14	0	N	14	2	42	0	9	2	47	0	N
87	4	46	5	3	7	41	8	2	N	2	48	4	4	8	42	6	2	S
88	5	52	1	0	12	43	2	1	S	3	55	0	0	16	40	1	1	S
89	28	3	26	1	15	0	43	0	N	30	2	25	1	12	0	46	0	N
90	4	48	6	0	3	48	6	1	N	6	52	0	0	5	48	4	1	N

* "Sig." refers to significance (at the 10% level) of differences in the proportion of "+"'s between high and low scoring groups; S = significant; N = non-significant.

lation between the Learning Ability and the Clerical Aptitude tests the highest scoring groups on the two tests included many of the same individuals, as did the lowest scoring groups. It was found that 25 items differentiated significantly between the high scoring and the low scoring Clerical Aptitude groups, adhering to the same criterion of significance as above, viz., P less than .10. These items are indicated in Table 1.

It will be observed from Table 1 that of the 90 statements covered in the Check List, 28 were found to be significantly related to scores on one or both of the tests used at the time of employment and hence are to some extent predictable through the use of these tests. Nine items are significantly related to scores on both tests; three are predictable by the Learning Ability Test but not by the Clerical Aptitude Test, and sixteen are predictable by the Clerical Aptitude but not by the Learning Ability Test.

Discussion

In order to see if any generalizations could be drawn about the nature of predictable items, as compared to those not predicted, the items were grouped as well as possible into several roughly homogeneous categories. While some items fell readily into the groupings, many statements were difficult or impossible to classify. Some seemed to fit almost equally well into more than one category. An attempt has been made to give a name to each group but the specific items themselves should be consulted to understand what is implied. From a study of the items as thus arranged, the following findings were derived:

(1) *Understanding of the work* (Items 20, 33, 36, 40, 48, 64, 71, 81, 88): Of nine items in the List which dealt with this subject, eight were predicted by the aptitude tests.

(2) *Errors in performance* (Items 2, 6, 9, 10, 13, 15, 30, 42, 44, 47, 49, 54, 55, 69, 70, 77, 79, 90): Of nineteen items dealing with the accuracy of the work, only three were predicted.

(3) *Quantity and speed of work* (Items 11, 23, 29, 43, 57, 63, 66, 68, 78, 82): All ten items in this group were predicted except Item 11, which was marked "True" for practically everyone, and Item 68, which deals with a special skill (note-taking).

(4) *Performance of multiple tasks* (Items 38, 65): Both items are concerned with the ability to handle effectively more than one task and both were predicted.

(5) *Unnecessary duplication in work effort* (Items 34, 87): Of the two items in this group, both were predicted by the Clerical Aptitude Test.

(6) *Typing* (Items 17, 31, 37, 50, 62, 79, 89): No item dealing with the quality of typing work was predicted.

(7) *Shorthand* (Items 45, 59, 68, 72): No item was predicted.

(8) *Grammar and spelling* (Items 16, 27, 46, 61): No item was predicted.

(9) *Statistical work* (Items 3, 21, 56, 75): No item was predicted.

(10) *Checking of one's work* (Items 8, 51, 73): No item was predicted.

(11) *Orderliness* (Items 1, 5, 35, 58, 74): No item was predicted.

(12) *Attitudes toward work and "personality" traits* (Items 25, 26, 28, 32, 60, 85, 86): No item was predicted.

Summary

In a comparison of ratings on a 90-item Check List for Clerical Workers with scores on an intelligence test and a clerical aptitude test, the results indicated several aspects of clerical work which are "predictable" on the basis of these tests, but pointed on the other hand to the need for separate tests in typing, shorthand, statistics, grammar, and spelling for positions in which these abilities are required. Even with this more complete battery, it is probable that only a portion of the total variance in on-the-job performance would be predicted, for it is not assumed that the Check List covered adequately all aspects of performance. As might be anticipated, the aptitude tests did not predict any of the so-called "personality factors." The Clerical Aptitude Test was efficient in predicting speed, amount, and understanding of work but was deficient in predicting accuracy. This may be due to the nature of this particular test, which is composed of a series of short subtests, and places a premium on quick grasp of directions and rapid work. Few errors are made on the test, and other studies have indicated that these errors have little significance in predicting on-the-job performance.

Received December 20, 1948.

A Study on the Use of a Work Sample

Marion Steel, Benjamin Balinsky, and Hazel Lang

Vocational Advisory Service, New York City

The O'Rourke "Ringing an Electric Bell" work sample, among many other work samples, has been used experimentally as a device for arousing and developing interest in various occupations.[1] About three years ago a number of Dr. O'Rourke's work samples were used in the NYA both to give short work experiences in various trades and to provide a measure of suitability for training in a specific trade.

The Vocational Advisory Service tentatively included the "Ringing an Electric Bell" work sample as part of its psychological testing program about one year ago and set about to evaluate it at the same time. The "Ringing an Electric Bell" work sample was chosen because an unpublished report stated that the time scores of the electrical work samples were found to have product-moment coefficients of correlation of about .50 with the ratings by foremen on the work of trainees in mechanical jobs such as machine shop. Other samples, such as one involving woodworking, gave much lower correlations. The "Ringing an Electric Bell" work sample took a short time to administer, was cheap in cost, using less expendable material, and was easy to set up for the next client.[2]

A try-out of the work sample, preliminary to the actual experiment, indicated that dexterities were involved in wiring the unit and also that the experience of those working on the sample was a factor to be considered. Remarks by those taking the work sample also indicated that the material was interesting. With these observations in mind, this study was designed to throw some light on the dexterities involved in the work sample, the effect of experience and the relative degree of interest shown in this material, in comparison with usual tests.

---

[1] Bulletin: Institute of Educational Research, Teachers' College, Columbia University, and the Civics Research Institute, Washington, D. C., Feb. 27, 1940. Kitson, Harry D.: Creating vocational interests, *Occupations*, 1942, 20, 567–571.

[2] The work sample was actually found to be practical for time and cost, the average time taken to complete the sample being 13 minutes and 40 seconds. It can be stopped at the end of 25 minutes since only 8% of all the subjects were still working at that time. The only expendable material was the wire and this cost about 1¢ per person.

Description of Sampling and Tests

For a period of two months the "Ringing an Electric Bell" work sample was given to a randomly selected group of clients who came to the Vocational Advisory Service for vocational guidance. Selection was limited only to the extent that the person was at least an elementary school graduate and between the ages of sixteen and twenty-five. This age group represents the bulk of the Vocational Advisory Service clients who come for guidance. Clients were also given some of the battery of tests ordinarily administered in vocational guidance.

At the end of two months, the sampling consisted of 86 individuals, 49 males and 37 females. The median age for the total group was 18 years and 9 months. The median educational attainment was high school graduation. Eighty-four per cent of the males and eighty per cent of the females had completed either some portion of the high school course or were high school graduates.

In administering the work sample, the following materials were laid out before the subject in a standardized manner: an electric bell, a pushbutton, three feet of insulated wire, a number 6 dry cell, a penknife, a pair of cutting pliers, a screwdriver, and a ruler. The subject was also provided with a sheet of instructions giving detailed step-by-step directions as well as diagrams.[1] Each subject was examined individually. The work sample was introduced by the examiner as follows: "This is a job in electricity we would like you to try. You do not need to have experience with this kind of material. Just follow the directions on the sheet. Start right here. (Point) You have all the material you need. Work as quickly as you can and be sure to follow the directions."

The total time was taken, including the time spent in reading directions. Speed was not emphasized but the stop-watch was plainly visible so that the subjects might be aware that they were being timed. The final instruction was to press the pushbutton and the work was considered complete when the bell rang. The work was stopped if it was still incomplete after thirty minutes.

The regular test battery included the following dexterity tests: the O'Connor Finger and Tweezer Dexterity tests and the Minnesota Rate of Manipulation, Placing and Turning tests. On the Finger Dexterity test, in accordance with the practice of the Vocational Advisory Service, the total time for the entire board was used as the score rather than the score specified by O'Connor. The scoring methods of the authors of the tests were used for the others.

---

[1] See original directions and diagrams by L. J. O'Rourke, Civics Research Institute, 3506 Patterson Street, N.W., Washington, D. C.

Experimental Procedure

The clients were divided into two groups. In group A (forty-three subjects), the above-mentioned four dexterity tests were given before the work sample and in group B (forty-three subjects), the order was reversed. The four dexterity tests were never given to more than two subjects at one time and the work sample was always given individually.

The work sample was administered individually in order to allow for careful observation of the handling of the tools and materials and the approach to the task. The examiners carefully recorded the use of directions, facility in handling the tools and material, initial adjustment to the task and the reaction to difficulties as well as spontaneous remarks. This was done in order to investigate the possibilities of establishing a qualitative rating. Each subject was interviewed briefly after the tests and the work sample had been administered. The following questions were asked: 1. What have you done before that was like this?; 2. Have you done any shopwork in school, repair work around the house, or a car?; 3. Which of these tests did you like best? Why?; 4. Did you try harder on one than another? Which one? Why?; and 5. Which was the easiest for you? Why?

Results and Interpretation

Pearson product-moment correlations were calculated separately between the scores on each dexterity test and the time score on the work sample. The Minnesota Spatial Relations test and the O'Rourke Vocabulary test had also been given as part of the battery of tests and Pearson product-moment correlations were computed between each of these tests and time taken to do the work sample.

The figures in Table 1 indicate that the correlations are low and that many of them are unreliable. Correlations between dexterity tests usually run somewhat higher, in the order of .40.[4] For the males, the correlation coefficients between the dexterity tests and the work sample are very low. This might be interpreted as meaning that the work sample is not measuring the same functions as the dexterity tests in the case of the males. The correlations for the females, although low, are higher than for the males and approach the order of .40. The differences between the correlations for males and females are significant in only the case of the placing test. These results must be interpreted in the light of the relatively small samplings.

The number of male and female cases are too small to make any definitive statements about sex differences but the boys made significantly

---

[4] Blum, M., and Candee, B., The selection of department store packers and wrappers with the aid of certain psychological tests. *J. appl. Psychol.*, 1941, 25, 76-85.

Table 1

Correlations of Work Sample with Various Tests

Test	Sex	N.	r	P.E. r	Significance of Correlation Differences Between Sexes
Finger Dexterity	M	49	.077	.095	
	F	35	.310	.090	2.00
	Both	84	.153	.072	
Tweezer Dexterity	M	49	.176	.093	
	F	35	.419	.094	1.83
	Both	84	.293	.067	
Placing	M	49	−.010	.096	
	F	35	.498	.080	4.0
	Both	84	.224	.070	
Turning	M	49	.097	.095	
	F	35	.347	.101	1.8
	Both	84	.103	.071	
Minnesota Spatial *Relations*	M	49	.247	.091	
	F	35	.304	.090	1.1
	Both	84	.260	.071	
Vocabulary	M	48	.220	.092	
	F	35	.317	.103	0.65
	Both	83	.204	.074	

higher scores on the work sample than the girls and this did not occur on any other test, as can be seen from an examination of Table 2. The critical ratio was 4.15 for the work sample and below 1.0 for each of the other tests given.

In order to test the effect of differences in experience between the males and females, data on the background of each subject were compiled. The data were available from school records, as well as from work histories and accounts of hobbies already obtained from the individuals by the counselors. The replies to the questions, "What have you done before that is like this?" and "Have you done any shopwork in school, repair work around the house, or a car?" asked at the close of testing, also gave information about experience.

The degree of experience was rated as none, little or some. Those rated as "none" had no experience at all or only woodworking in elementary school. The latter was the base experience for the males. "Little" experience was defined as occasional minor repairs at home or as a hobby, usually referred to as "fixing things"; or two shop courses, not including electricity; or a short time (approximately three months) experience in a factory. The criteria for "some" experience were two or more shop courses other than woodworking; much and more extensive home or

Table 2

Means, Standard Deviations and Critical Ratios Between Sexes

Test	Sex	N.	M.	S.D.	C.R.
Finger Dexterity	M	49	8.00	1.66	
	F	35	9.03	1.67	0.104
	Both	84	8.95 min.	1.41	
Tweezer Dexterity	M	49	6.15	0.93	
	F	35	6.18	1.11	0.132
	Both	84	6.17 min.	0.91	
Placing	M	49	239.18	21.12	
	F	35	242.20	20.15	0.683
	Both	84	240.48 sec.	20.40	
Turning	M	49	190.10	26.20	
	F	35	192.14	21.19	0.305
	Both	84	190.52 sec.	22.65	
Minnesota Spatial Relations	M	49	9.96	2.18	
	F	35	9.80	1.82	0.800
	Both	84	9.90 min.	1.99	
Vocabulary	M	48	65.95	13.25	
	F	35	68.80	15.70	0.769
	Both	83	67.14	14.40	
Bell and Battery	M	49	11.70	4.56	
	F	35	16.64	5.76	4.15
	Both	84	13.81 min.	5.62	

hobby experience; or an electrical shop course and at least one other shop course; or finally three different shop courses, not including electricity.

Examination of Table 3 indicates a consistent trend for those with more experience to complete the work sample in less time. Apparently experience is reflected in the time scores. Blum and Candee, in the study referred to above, found that experience was a factor in raising the scores on tests requiring the handling of concrete materials, specifically the placing, turning, and finger dexterity tests. They wrote, "Apparently

Table 3

Amount of Experience, Median Time Scores, and Interquartile Ranges for Each Sex on Work Sample

Experience	Sex	No.	Median Time	Q
None	M	20	12' 30"	2' 20"
	F	24	18'	3' 45"
Little	M	8	10' 30"	4' 00"
	F	6	17'	3' 38"
Some	M	20	9' 50"	1' 48"
	F	5	9' 15"	38"

experience in wrapping does have a slight effect in raising test scores on three different tests and in reducing initial differences among the workers on the tests."

To test quantitatively the relative "interestingness" of the work sample, and the other tests, the question "Which of these tests did you like best?" was asked. The experimental groups had been divided into Group A and Group B. Group A had the dexterities first, Group B the work sample first. In Group A, 20 of the males and 11 of the females liked the work sample best; 2 of the males and 4 of the females liked one of the dexterities best, and 4 of the males and 1 of the females had no particular preference. One girl of the 17 girls in the A group refused to complete the work sample. In Group B 18 of the males and 12 of the females liked the work sample best, 3 of the males and 5 of the females liked one of the dexterities best, and 1 of the females had no particular preference. One girl of the 19 girls in Group B refused to complete the work sample. For the total male group, the standard error of the difference between the per cent liking the work sample best and the per cent not liking the work sample best was 7.7. This difference is significant. For the total female group, the standard error of the difference is 2.6. This would mean that the chances are about 9 in 1000 that the difference is a chance difference due to sampling.

Table 4

Number of Answers to Question, "Which of These Tests Did You Like Best?"

	Group A		Group B		Total	
Test Liked Best*	M	F	M	F	M	F
1. Work Sample	20	11	18	12	38	23
2. A Dexterity Test	2	4	3	5	5	9
3. No preference	4	1	0	1	4	2

* One female each in Groups A and B refused to complete the work sample.

Another question, "Which was easiest for you?" was also asked. In Group A, only 11 of the males and 4 of the females said that they found the work sample easier, 12 of the males and 12 of the females said that they found one of the dexterities easier, and 3 of the males and 1 of the females said no one test was easier than another. In Group B, only 8 of the males and 2 of the females indicated that the work sample was easier, 10 of the males and 12 of the females found one of the dexterities easier and 3 of the males and 5 of the females said no one test was easier. Evidently the greater degree of interest in the work sample was not due to the fact that it was easier than the other tests.

Some examples of replies to the question why the work sample was liked follow: "This is more like real work, but those are like child's play." "Just liked it. Get something out of it, the sound." "Made something, know it's finished. It works. Something I made." "Has more sense to it."

Table 5

Number of Answers to Question, "Which Was Easiest for You?"

	Group A		Group B		Total	
Easiest Test	M	F	M	F	M	F
1. Work Sample	11	4	8	2	19	6
2. A Dexterity Test	12	12	10	12	22	24
3. Neither Kind	3	1	3	5	6	6

Tentative Conclusions

The following conclusions can be drawn from this preliminary study:

1. The work sample had low correlations with the dexterity tests. This might be interpreted to mean that the work sample was measuring functions different from those measured by the Finger and Tweezer Dexterity tests and by the Placing and Turning tests.

2. A significant sex difference was obtained on the work sample for the group used in this study. This sex difference must be considered as preliminary and might possibly be attributed to differences in degree and kind of experiences had by the males and females in this sampling.

3. The amount of experience was related to the time taken to complete the work sample. Those with "some" experience completed the work unit in less time than those with "little" or "none."

4. The work sample was liked best by most individuals tested, both male and female, although a greater percentage found it more difficult than the dexterity tests.

Recommendations

Qualitative descriptions of the performance on the work sample gave valuable information to the vocational counselors. The descriptions were in terms of work habits and attitudes, facility in handling the tools and material, initial adjustment to the task and the reaction to difficulties. The examiners had checked each other for reliability of the qualitative description. However, in order to make the work sample more applicable to an industrial situation, it is thought necessary to have the qualitative descriptions checked by qualified people in industry engaged in such work as, for instance, that of electrical assembly.

The time score on the work sample also needs validation if the sample is to be related to success on a job. It is proposed that the work sample be tried out in several shops, such as radio or electrical assembly, where shop foreman ratings would be available. The ratings could then be compared with the time score to test the validity of the time score.

Received December 6, 1943.

A Method of Objectively Measuring Shop Performance *

Marion White McPherson

Wayne County Training School, Northville, Michigan

The need for some device for the diagnosis of trainability and for the refined evaluation of achievement in wood shop work is well known. A survey of the literature reveals few objective measures of performance in this area. In their study for the Committee on Human Migration the Minnesota group[1] used actual shop performance to determine the validity of their tests. The idea for the compilation of our test originated from the methods they used. It was possible that their technique could be developed into a short, convenient evaluation and thereby make their criteria for the validity of a test the actual performance to be measured. Therefore, we have begun an investigation of the practicality of such a direct measurement and of its sensitivity to continued wood shop experience. Refined data regarding the reliability and the prognostic value of this device, once developed, are matters for further research.

For the copying of a model wood block to constitute a satisfactory measure of wood shop achievement, it must be amenable to objective scoring and it must include as many as possible of the basic activities. To determine these we were assisted by our wood shop teachers[2] since they were able, out of their experience, to identify the important operations and the precision which our mentally defective population might be expected to achieve. Work with at least the saw, drill and chisel should be included. The product should be scored with respect to accuracy of dimensions, angles, and locations; to method of determination of positions of operation; and to neatness of execution. In addition, the convenience of the device would be increased could the sampling be integrated into a single piece of wood. All of these considerations entered into the construction of the model.

---

* From the Wayne County Training School, Robert H. Haskell, M.D., Medical Superintendent, Northville, Michigan. Studies in the Psychopathology of Childhood and Mental Deficiency, supported by a grant from the McGregor Fund, Detroit, Michigan. Report No. 65. The achievement measurements described in this paper were devised by A. A. Strauss, Z. P. Hoakley, and L. C. Sullivan.

[1] Paterson, D. G., and Elliott, R. M., et al., *Minnesota mechanical ability tests.* Minneapolis: The University of Minnesota Press, 1930.

[2] We wish to express our appreciation to Mr. Edmund Crosby and Mr. Norman Running.

In order to assure uniformity of approach the model was presented in four consecutive stages of completion. Each of the four blocks was 10 inches long, 5½ inches wide, and ¾ inch thick. The first block was a plain board, cut to certain dimensions; the second had the hole drilled in the left side; the third had the above and also the central groove; and the fourth was the completed block. These blocks, in a left to right progression, were hung on the wall in front of the subject but beyond his reach.

To insure objective evaluation, a scoring pattern was developed. This consisted of an outline of the model, drawn on a sheet of transparent plastic, that could be superimposed on the subject's completed block. In addition to the outline of the pattern, lines were drawn at uniform intervals to indicate the degrees of deviation of the product from the model. The appropriateness of the intervals was determined by the precision which, with training, these children might be expected to achieve. For example, lines were placed at each one-eighth inch interval for a reasonable distance at either side of the lines marking the correct length of the board. The line marking each correct position was assigned a definite value which was reduced by one point for each unit of deviation. That is, a block of the correct length was scored 10; one that was one unit (⅛″) either too short or too long scored 9; one that deviated in either direction by as much as two units (¼″) received a credit of 7, etc.

In an attempt to determine the feasibility of measuring and scoring performance with the method outlined in the preceding paragraphs, the model was presented to fifty-nine boys of the Wayne County Training School who were enrolled in the wood shop course in the academic year 1940–41. As enrollment in this shop is a part of the natural sequence of the training program for boys, there were no known extraneous factors operative in the selection of subjects. All were in their thirteenth year. The mean Binet P.C.[8] was 86.49, S.D. 6.31; the Arthur Performance P.C. 89.93, S.D. 6.78.

One boy at a time was admitted to an enclosure of approximately 10 by 40 feet which had been partitioned off at the end of our wood shop. Here he had access to a variety of tools including different sizes of chisels, saws, drills and, of course, the correct implements for the assigned task. He was given a board of the width and thickness of the models but at least one yard in length and was told to make his board look first like the one hanging in front of him at the extreme left, then like the second, then like the third, and finally like the one at the extreme right. No more specific instructions were given. He was allowed to work without time limit and with only casual adult supervision.

---

[8] Hildon, A., *Table of Heinis personal constant values.* Minneapolis: Educational Test Bureau, 1933.

One semester later fourteen of the subjects were again presented with the problem of reproducing the model. This was done in order to measure the change in score which would result from attendance in the wood shop for two hours per day during the half year.

When the subjects had completed their blocks, a technique for quantifying the performance was developed.[4] Each raw score was multiplied by certain numbers ranging from 1 to 6 depending upon the judged difficulty of the task and the number of times the particular operation was scored. Neatness and method of determination of position of operation were evaluated on a point scale. The number of intervals on the scale was determined by the units that gave the best approximation of a normal probability curve. The total possible score of 300, divided by 3, left a maximum score convenient to treat and sufficiently large to express fine gradations of skill.

That our scoring method is reliable is indicated by the results of a brief study. After eight months of no contact with this research the psychometrician, who originally scored the boards, rescored them. Although 82 readings are required on each of the 59 boards the Pearson product moment coefficient of correlation between the two scorings was +.97, ±.01.

Table 1

The Means and Standard Deviations of the Raw Scores on the Wood Shop Achievement Measurements Obtained by the Entire Group and the Training Group

	First Scoring			Second Scoring		
	Number of Boys	Mean	S.D.	Number of Boys	Mean	S.D.
Entire group	59	50.51	14.05	—	—	—
Training group	14	55.93*	15.52	14	70.02*	16.00

* Fisher's t for related measures = 4.42, significant at the 1% level.

Table 1 indicates the mean and standard deviation of the raw scores for the entire group and for the training subgroup. Table 2 presents in terms of frequency the changes in scores of the 14 subjects after a half year in the wood shop. Fisher's test of significance indicates that the mean gain of 15 points between scores on the first and second measures would be found less than one per cent of the time if the means did not differ significantly from zero.

---

[4] See manual and record blank for detailed description. Models, manuals, scoring patterns and record blanks for both wood and metal shop measurement may be purchased from C. H. Stoelting Company, 424 N. Homan Avenue, Chicago, Illinois.

A Pearson product moment coefficient of correlation between the raw scores for the 59 subjects and their Binet ratings was found to be +.43, significant at the one per cent level; a similar coefficient between these values and the Arthur ratings was found to be +.54, significant at the one per cent level.

Table 2

The Amount of Change in Scores Between the First Measurement and the One Following Training in the Wood Shop (14 individuals)

Amount of Change	Frequency
−11 to − 7	1*
− 6 to − 2	1*
− 1 to + 3	0
+ 4 to + 8	2
+ 9 to +13	3
+14 to +18	0
+19 to +23	2
+24 to +28	3
+29 to +33	2
	T = 14
	Mean: +15

---

* Considering the instability of a number of our children the two training score decrements are not unexpected.

The distribution of the raw scores and the retest gains for our group indicate that the technique has value. The application of this measure to a normal or superior population may demand changes in the precision units and the establishment of norms suitable to that group. Thorough investigation of this technique involves control of academic subject achievement, pre-school shop experience in wood work, verbal or non-verbal superiority, race, bilingualism, etc.

Measurements for Metal-Shop

The wood-shop study has indicated that an activity can be measured directly without the necessity of evaluating performance through the use of tests that merely sample behavior. There is no *a priori* reason why such a technique could not be extended to meet the needs of other shop activities, for example, those of a metal shop.

To investigate this, another staff conference was held. The important activities in the metal shop were isolated, their relative difficulty determined, and the precision which our children might be expected to attain identified. Inspection of the work sample evaluated in the Minnesota study (1) did not reveal any one product that would sample as many activities as we desired. Consequently, we developed four patterns, the reproduction of which necessitates wire bending, sheet metal soldering, riveting, locked seaming, wiring, and circular and angular cutting.

Thirty-three 14-year-old boys were asked to reproduce the model. Only the tools necessary for the task were present and all the patterns were accessible throughout the reproduction. In this situation the measurement is one of the efficiency of the use of tools and does not involve their selection. No specific instructions were given to the subjects.

The scoring pattern is of the same type as that used in the wood-shop study. It can be superimposed over the bent wire, the cut pattern, and the folded metal. Although we have been able to give but superficial study to this device, we have compiled the manual and scoring key as a means of presenting in detail the important operations that are amenable to objective rating.

Received February 11, 1944.

An Analysis of Absenteeism in One War Plant

Neal G. Schenet

Elgin National Watch Company, Elgin, Illinois

To say that absenteeism is of prime importance in industry today and that it is one of the largest problems to be faced on the home front is redundant. The mere fact that the personnel man in industry now is confronted with the subject wherever he turns, in articles and on the air, as well as in the plant, is proof enough.

It will be the purpose of the present study to determine the nature and extent of absenteeism in a typical war production plant with special reference to the individual and the collective effects of age, sex, and length of company service. Sex differentials are always an obvious starting point, and a review of the literature on the subject shows that age and length of company service would be of interest also. Almost any other differential could have been used, such as physical characteristics, intelligence test scores, etc., but these appear, at least on the surface, to have little relationship to the total problem. This led to the choice of the variables used in the present study.

Review of the Literature

In reviewing the available literature on absenteeism, it was found to fall generally into three classes; namely, data on causes, discussion of records and methods of study, and suggestions as to remedies.

Most of the articles have been unscientific in their approach, omitting all or nearly all statistics on the subject, making only broad generalizations, and usually being on a speculative, rather than on a practical level. Estimates, in these writings, as to the number of hours lost, types of absences, etc., vary widely. Bearing this out, one governmental source states, "There is no statistical information available to indicate the general extent of absenteeism in the war industries. Scattered reports from a number of factories reflect rates ranging from between two and three per cent to fifteen per cent or more" (11:2). Certain general characteristics of absenteeism, as given by several governmental reports, may be mentioned. One source states that ". . . absenteeism rates are generally higher for women than for men, even on jobs of the same general character. . . . Greater sickness rates among women are probably a factor in their higher absence rates . . ." (11:2).

On a basis of comparison of male and female rates we learn, "One large war plant reports current absence rates at 4.8 per cent for men and 7.4 per cent for women. . . . These figures are typical of a number of reports" (11:2). Age may be a factor of importance since "a study undertaken by one company indicates that absenteeism tends to be higher among older workers, increasing rapidly after forty or fifty years of age" (11:3).

There seems to be a series of effects which cause a tendency for absences to be numerous on days adjacent to a week-end or holiday. "These effects frequently combine to produce . . . the highest (rate of the week) on Saturday" (11:3).

Absence figures appear to indicate that offices, tool cribs, and supervision show lower rates than factory work generally. There is, however, no information available, to the writer's knowledge, to indicate whether absences tend to be relatively more frequent on routine as compared with nonroutine work, or on heavy versus light work.

The definition of absenteeism suggested by the United States Department of Labor, and used throughout this paper, is as follows: "Absenteeism is the absence of a worker during a full shift that he is scheduled to work" (11:1).

In conclusion, as far as the writer could determine from going through the literature, very little work has been done in the field of causes of absenteeism, or more specifically, in the field of the effects of certain variables upon the total field of absences and absenteeism.

Materials and Methods

In order to make a meaningful and scientific survey of the problem and yet to keep it within reasonable limits as to size and scope, it was decided to break the absence figures down by (1) age groups, (2) length of company service groups, and (3) sex groups.

For the purposes of this investigation it was decided that one of the plants of the Elgin National Watch Company, by which the writer is employed, be used because of comparative ease of access to absence records, familiarity of the writer with the plant and its personnel, and size of the plant. The plant chosen was Company Plant No. 2, manufacturing a mechanical time fuze for the armed forces. This work is of a fine, precision nature, requiring in general a higher type of employee than the average factory. The plant has an average labor force of approximately 850 to 900, with about 65% women and 35% men. The regulation work week at present is six eight-hour days, with Saturday an overtime day and not, in the past history of the company, normally a working day.

The period of study used was the first four months of 1943: January, February, March, and April. The days of the week studied were Monday, Tuesday, Wednesday, Thursday and Friday. Saturday was omitted specifically because of the fact that absences on this day become of importance in and of themselves for certain very definite reasons. For example, Saturday is the last day of the week and fatigue will enter into the situation; also, Saturday is normally an overtime day and as such is considered as "different" by the average employee. The writer found, merely by inspection, that these statements are true in this plant as shown by the unusually high number of absences for this day in proportion to all others. It was felt, as a result of these observations, that Saturdays should be removed from the study, and if surveyed at all, should be the subject of a further, separate study.

The age groups chosen were (1) thirty years of age and under, (2) between thirty and forty years of age, and (3) over forty years of age. The length of company service groups chosen were (1) three months and under, (2) three to six months, and (3) over six months. Subsequently, it was found that the length of company service groups were somewhat out of line, being weighted on the side of longest service. Thus, the probable error of results found in some of the service groups will be larger than if the cases had been more evenly divided. However, in spite of this skewed distribution, the facts obtaining in the small groups appear to corroborate reasonably well those in the larger groups.

In determining which statistics to use, it was decided to use the number of days lost and also, to disregard any absences under one day in length. This last decision was made because of the labor of handling the data and because, upon inspection, the basic facts appeared to hold true regardless of the length of the absence. "It is . . . difficult (and usually unnecessary) to tabulate part-day absences and there is, for example, no obvious line of demarcation between part-day absenteeism and tardiness" (11:1).

Individuals used as subjects in the survey were classified as to age and length of company service at the beginning of the period; that is, on January 1, 1943. Because of a possible error introduced by employees entering and leaving the service of the company during the period, it was determined to use only those individuals who remained active in one specific department of the plant during the course of the study. The total number of subjects obtained in this manner was 750; 280 men and 470 women.

The general method used in obtaining the raw data was to use the figures on the company's "Daily Time Exception Sheet," which is made out by each department and turned in to the Payroll Department daily.

On these sheets, only those absences listed as "S" (Sickness), "WP" (Without Permission), and "TR" (Time Requested) were used, all others (Plant Accident, Vacation, etc.) being omitted as not pertinent to the survey. In compiling these statistics and throughout the study, WP and TR have been combined into one group which will be referred to as "P" (Personal).

A brief word of explanation for this procedure is in order here. Because of the method of marking absentees in this plant, it is felt that both WP and TR actually contain a great deal of each other and should not, under the present conditions, be separated for the purposes of this study. Absences are reported either by telephone, by a friend of the absentee, or in person when the absentee returns to work. As a result of this, if a person called his department and stated, "I will be absent today because I am going to see the doctor," one department may class it as WP because no prior permission was obtained, while another department may class it as TR merely because the person was kind enough to telephone and not leave them in doubt. This procedure has been verified by the writer in personal conversations with the various department heads. Since we cannot separate these intangible amounts of WP in the TR heading it appears more logical to group them together.

No attempt was made to show the duration of the absences, each full-day absence being listed as a separate item regardless of whether or not it appeared in connection with other full-day absences. That is, while we may know how many full days an individual was absent during the calendar month, it was not recorded whether or not those days absent were scattered throughout the month or localized in one long absence.

When this master list of absences was completed it was possible to obtain the number of days lost for sickness and personal reasons by department, sex, age, length of service, or by any one or more of these variables without regard for the others. Distribution of the data by these groups was facilitated by entering the pertinent facts upon a series of index cards, one for each employee in the survey. These cards then contained (1) the name and department of the employee, (2) the age group of the employee, (3) the length of company service group of the employee, and (4) all absences listed for the employee broken down by calendar month as well as reason for absence. These cards were then distributed by the groups listed above.

The absence figures were then used to calculate the absence rates, by means of the following formula: $\text{Rate} = \frac{\text{Number of days lost by group}}{\text{Number of persons in group}}$.

By means of this calculation it is now possible to compare any group with any other group or combinations of groups, since we are dealing in rates rather than raw data. Also, in order to provide a basis for comparison with other sets of statistics, both national and local, the absenteeism rates by men and women for the entire four month period were figured through the use of the following formula, suggested and used by the Bureau of Labor Statistics:

$$\text{Absenteeism Rate} = \frac{\text{Man-days lost} \times 100}{\text{Man-days scheduled to work}} \text{ where}$$

Man-days scheduled to work = Number of persons in group × 86, the number of work days involved in the study.

There may be some question as to the use of two different formulas for computing absence rates. The reason for using the first formula is to simplify the labor involved, since the second formula requires an additional computation. It is not necessary to have the rates computed by the Bureau of Labor Statistics formula in all the individual groupings, since there is no method of checking and comparing such figures with national or area rates. There obviously is a one-to-one correspondence between the writer's formula multiplied by 100/86 and the Bureau of Labor Statistics formula.

Upon inspection of the data, several differences appeared to be outstanding, and in order to determine whether they were significant and real differences or whether they might easily have occurred due to chance, standard deviations of certain of these items were computed. From these, critical ratios were determined, and results will be discussed later in this paper.

Results and Discussion

In a discussion of the results of the present study, it would be most pertinent to begin with the field of sex differentials in total and proceed from there to the differences within each group and the age and service differences which appear to stand out upon inspection of the data.

The most striking fact is that the female rates are proportionately much greater in almost every group than the male rates. This appears true throughout the entire Plant, in all departments, service groups, age groups, and totals. In total absences in all departments the female rate is exactly three times the male rate, that for men being 1.3 while that for women is 3.9. For sickness absences the rate for women is twice the male rate (1.9 and 0.7). In personal absences the difference is even greater, the female rate being between three and four times that of the men (2.0 and 0.6). (It will be recalled that all rates may be thought of as "average number of days lost per employee.") While these results

Table 1

Absence Rates for Male Employees in Four Largest Departments and for Total of All Departments Included in the Survey

		Service Group 1				Service Group 2				Service Group 3				Total			
Department	Age Group	No. of Cases	S	P	T	No. of Cases	S	P	T	No. of Cases	S	P	T	No. of Cases	S	P	T
Final Assembly	1	0	0.0	0.0	0.0	0	0.0	0.0	0.0	7	1.5	0.7	2.2	7	1.5	0.7	2.2
	2	0	0.0	0.0	0.0	0	0.0	0.0	0.0	6	1.1	0.6	1.8	6	1.1	0.6	1.8
	3	1	0.0	0.0	0.0	2	0.0	0.0	0.0	5	0.4	0.0	0.4	8	0.4	0.0	0.4
	Total	1	0.0	0.0	0.0	2	0.0	0.0	0.0	18	1.1	0.5	1.6	21	1.1	0.5	1.6
Automatics	1	0	0.0	0.0	0.0	4	0.0	0.5	0.5	21	1.0	1.1	2.1	25	0.9	1.0	1.8
	2	2	0.0	0.5	0.5	4	0.0	0.5	0.5	29	0.8	0.6	1.4	35	0.7	0.6	1.2
	3	1	0.0	0.0	0.0	4	0.0	0.5	0.5	14	0.5	0.2	0.8	19	0.4	0.3	0.7
	Total	3	0.0	0.3	0.3	12	0.0	0.5	0.5	64	0.8	0.7	1.5	79	0.7	0.7	1.3
Plate	1	1	0.0	0.0	0.0	4	0.5	0.2	0.7	14	0.6	1.0	1.5	19	0.5	0.7	1.3
	2	1	0.0	0.0	0.0	4	0.2	2.5	2.7	17	0.2	0.4	0.6	22	0.2	0.7	1.0
	3	3	0.0	0.6	0.6	4	0.2	2.2	2.5	18	0.6	0.2	0.8	25	0.5	0.6	1.1
	Total	5	0.0	0.4	0.4	12	0.3	1.6	2.0	49	0.5	0.5	0.9	66	0.4	0.7	1.1
Sub-Assembly	1	1	0.0	0.0	0.0	0	0.0	0.0	0.0	14	0.6	0.2	0.9	15	0.6	0.2	0.8
	2	0	0.0	0.0	0.0	1	1.0	1.0	2.0	11	0.0	0.1	0.1	12	0.1	0.3	0.3
	3	0	0.0	0.0	0.0	3	0.0	0.0	0.0	9	0.5	1.6	2.2	12	0.4	1.2	1.7
	Total	1	0.0	0.0	0.0	4	0.3	0.3	0.5	34	0.4	0.6	1.0	39	0.4	0.6	0.9
All Departments	1	3	0.0	0.0	0.0	10	0.3	0.3	0.6	73	0.9	0.7	1.6	86	0.8	0.6	1.4
	2	6	0.6	0.8	1.5	12	0.2	1.1	1.3	75	0.5	0.5	0.9	93	0.5	0.6	1.1
	3	8	0.0	0.6	0.6	16	0.1	0.8	0.8	77	1.1	0.5	1.6	101	0.8	0.6	1.4
	Total	17	0.2	0.6	0.8	38	0.2	0.7	0.9	225	0.8	0.6	1.4	280	0.7	0.6	1.3

Code: S = Sickness; P = Personal; T = Total

Table 2

Absence Rates for Female Employees in Four Largest Departments and for Total of All Departments Included in the Survey

Department	Age Group	Service Group 1				Service Group 2				Service Group 3				Total			
		No. of Cases	S	P	T	No. of Cases	S	P	T	No. of Cases	S	P	T	No. of Cases	S	P	T
Final Assembly	1	0	0.0	0.0	0.0	22	2.7	3.5	6.2	76	1.9	1.0	2.9	98	2.1	1.6	3.7
	2	1	11.0	0.0	11.0	10	0.7	1.9	2.6	24	2.9	3.2	6.2	35	2.5	2.7	5.3
	3	2	0.0	0.5	0.5	7	2.1	3.0	5.1	19	3.4	3.2	6.7	28	2.8	2.9	5.8
	Total	3	3.7	0.3	4.0	39	2.1	3.0	5.1	119	2.3	1.8	4.2	161	2.4	2.1	4.4
Automatics	1	0	0.0	0.0	0.0	0	0.0	0.0	0.0	4	2.0	1.0	3.0	4	2.0	1.0	3.0
	2	0	0.0	0.0	0.0	1	0.0	1.0	1.0	6	0.3	1.0	1.3	7	0.3	1.0	1.2
	3	1	0.0	2.0	2.0	1	7.0	0.0	7.0	1	0.0	0.0	0.0	3	2.3	0.6	3.0
	Total	1	0.0	2.0	2.0	2	3.5	0.5	4.0	11	0.9	0.9	1.8	14	1.2	0.9	2.1
Plate	1	1	21.0	6.0	27.0	5	0.4	3.8	4.2	33	1.7	2.6	4.3	39	2.1	2.9	4.9
	2	1	0.0	1.0	1.0	2	1.5	0.5	2.0	24	1.1	2.3	3.5	27	1.1	2.1	3.3
	3	7	0.0	0.4	0.4	11	4.0	2.7	6.7	39	1.5	1.7	3.3	57	1.7	1.7	3.6
	Total	9	2.3	1.1	3.4	18	2.7	2.7	5.5	96	1.5	2.2	3.6	123	1.7	4.2	3.9
Sub-Assembly	1	2	0.0	4.0	4.0	2	0.0	0.0	0.0	55	1.6	1.6	3.4	59	1.5	1.7	3.3
	2	3	0.0	1.3	1.3	6	1.6	3.0	4.6	40	1.6	2.2	3.8	49	1.5	2.2	3.8
	3	9	1.7	2.0	3.7	10	1.1	2.2	3.3	34	1.5	2.2	3.7	53	1.5	2.2	3.6
	Total	14	1.1	2.1	3.2	18	1.1	2.2	3.6	129	1.6	1.9	3.6	161	1.5	2.0	3.6
All Departments	1	3	7.0	4.7	11.6	29	2.1	3.3	5.4	173	1.9	1.5	3.4	205	2.0	1.8	3.8
	2	6	2.2	0.8	3.0	19	1.1	2.1	3.1	97	1.7	2.4	4.1	122	1.6	2.3	3.8
	3	19	0.8	1.3	2.2	30	2.9	2.6	5.5	94	1.8	2.2	4.1	143	1.9	2.1	4.1
	Total	28	1.7	1.5	3.3	78	2.2	2.7	4.9	364	1.8	1.9	3.8	470	1.9	2.0	3.9

Code: S = Sickness; P = Personal; T = Total are in line generally with those obtained in other studies it would seem that the differences are even more pronounced.

A statistical investigation of these results yields, in the case of male totals versus female totals, a critical ratio of 8.6, indicating a very high probability that the difference is a real and significant one. In the case of male, sickness, versus female, sickness, the critical ratio is 5.2, and it is 9.0 for male, personal, versus female, personal.

One other fact of interest is that for the total group the sickness and personal absences are about evenly divided, the former accounting for 49% of the total and the latter for 51%. This is at variance with a statement made by Spriegel and Schulz that "In a study of 10,000 employees of both sexes it was found that the causes of their absences were 40% due to sickness and accident and 60% due to personal reasons" (10:164). However, this may also be due to a difference in definitions.

In the field solely of male rates, certain factors appear to be of interest. The rates for sickness and personal absences are approximately the same, being 0.7 and 0.6, respectively, showing a fairly even distribution of absences on this basis. Of the length of company service groups, Group Three, or Over Six Months, appears to be the highest, especially from the standpoint of sickness absences. This would have no apparent explanation other than the possible fact that as length of service increases, more adherence to the rules of reporting absences follows, which would lead to classifying many absences as sickness which would otherwise be classed as personal absences, because of not being reported. The critical ratio here for men with more than six months service versus men with zero to six months service (groups combined because of the small number of cases in each) is 4.1.

The age groups of the men do not appear to be related significantly to absence rates. Group Two, thirty to forty years of age, is slightly lower in sickness absence rates, but not appreciably so. This seems to be at variance with the statement made by Watkins and Dodd that, ". . . the time lost by male workers below the age of forty on account of illness tends to be lower than the average male disability, but beyond forty, males show a rapidly increasing morbidity rate" (15:266). They also state elsewhere, "Experience indicates that youthful employees are more careless in the matter of punctuality and attendance than are more mature workers" (15:265). While this may be true in general, the present study does not seem to show this to any great degree in this particular plant. The fact that these findings do not agree entirely with the results of other studies may be due to certain special factors obtaining in this specific plant. However, it is evident that the results are showing a general trend toward agreement, since both the younger and older workers are slightly higher in rate than the intermediate group, as far as sickness absences are concerned.

In the field of female rates, it is apparent that here too, the sickness and personal absences are approximately the same, the rates being 1.9 and 2.0 respectively. Here, however, Service Group Two, three to six months, is highest in rate, being 1.1 higher than the next group and 1.6 higher than the lowest. There appears to be no logical explanation for this unless, as the writer feels, there is a period of orientation during which the new worker tends to be quite regular in attendance, followed by a period of laxity in reporting absences, caused, perhaps, by increasing familiarity with the plant, after which there is a lapse into a somewhat steady groove of cooperation. From the statistical standpoint, this difference is not as significant as some others, for the critical ratio in the case of women with three to six months service versus women with over six months service is only 1.4. However, even this critical ratio indicates a probability of .92 that the difference is real.

Age group one, 30 and under, is no higher than the other groups, somewhat reversing the findings of Watkins and Dodd from the standpoint of the increased absenteeism of younger workers. However, age group three, over 40, is slightly higher than the other groups, bearing out their statement, "In the case of female employees the rate remains less than the average up to age thirty, but increases beyond that point" (15:266). In the female rates, the service groups show more variation than do the age groups, as in the case of the male rates, tending to point to the fact that length of service is more important than age.

In the realm of departmental differences, the outstanding fact is that Final Assembly has more absenteeism than any other department, of the four largest ones. (In discussing departmental differences, only the largest departments will be used, since they contain nearly 89% of the total number of cases.) This holds true in sickness and total absence rates, and is evident in the rates of both men and women employees, although in the case of men the rate is not markedly higher than in the other departments. In this department the sickness rate of both male and female is much higher than in the other departments, whereas rates of personal absence are only higher in the case of women, and then just moderately so. In the case of men, sickness, Final Assembly versus men, sickness, Sub-Assembly, the rates are 1.1 and 0.4 respectively, with a critical ratio of 1.2, indicating 88 chances in 100 that the difference is real. For women in the same comparison, the rates are 2.4 and 1.5 respectively, with a critical ratio of 2.4, indicating 99 chances in 100 that the difference is real. Thus, the difference for women is more significant than for men.

Upon first glance this may seem as though there is relatively more sickness in this department than in others. This is undoubtedly true, but along with this, the personal absences, at least for women, are also considerably higher than in the other departments. This particular department, having more trouble with absenteeism than the others, has been conducting intensive educational campaigns on the subject. Their main emphasis has been upon the need for employees to report their absences whenever possible. Also, strong punitive measures have been instituted in this department, so that, for example, three unexcused absences bring about discharge of the absentee. However, as shown by the absence rates, these practices have not greatly reduced the total problem.

It has been the general hypothesis in this department that the younger workers caused the majority of the absences here, due to sickness, irresponsibility, etc. The rate charts do not bear this out, however, and, as a matter of fact, they point out that the younger persons in this department (particularly the women) tend to have *lower* rates than do the older employees.

It is extremely interesting, as well as puzzling, to note that the absence rates in general for Final Assembly are appreciably higher than those for other departments, in spite of the educational work on the subject and the disciplinary measures in effect. There is nothing apparent in the attitude of supervision here which would tend to influence the absence rates, and the psychological factor of working on a finished product rather than a small, perhaps unrelated part, should have a constructive influence also. Since the employees must report their absences more carefully, it is possible that there is a tendency for the sickness absence figures to increase, for it is much easier for an employee to state that he is ill than to state that he wants the day off for shopping, or a trip. However, this does not explain the increased personal absences and the resultant increased totals, and there is nothing in the study which would give the answer to the unusual situation obtaining here.

Further study of departmental differences indicates that in both male and female, the Plate Department has a higher personal absence rate, appreciably so in the case of women. For example, in the case of women, Plate, personal absences versus women, Sub-Assembly, personal absences, the rates are 4.2 and 2.0 respectively, with a critical ratio of 5.6. Such a difference is not evident anywhere else in the departmental statistics. Two factors inherent in the department itself have probably combined to cause this difference. First, the type of work in this department is large, rough and dirty as compared to that in the other departments, and secondly, the production schedules in this department have been such that there were peaks and valleys in the amount of work available to the employees. Selection of employees for this department generally led to placing less skilled labor there, and such persons tended to be rather irresponsible in their attendance records and in reporting their absences. Also, since production varied so greatly, a certain amount of resentment and lack of interest on the part of the employees was detectible, leading to higher absenteeism. During the period of the survey, a temporary lay-off was in progress in this department, and while these figures were not included in this study, the psychological effect upon the remaining employees, from the standpoint of their apparent need to the plant and their job security, may have been reflected in their attendance records.

A brief word is in order from the standpoint of the total plant as compared to other plants in the same area and to national figures. In general, the absence rates are very low compared to other similar plants, perhaps due to a concentrated program against absenteeism and the generally higher type of employees, on the whole, as a result of the precision quality of the work. The following table gives the pertinent facts:

Table 3

Percentage Absence Rates, April 1943, and Percentage Women Among Wage Earners, April 1943, in Selected Industries *

Industry	Absence Rate	Per cent of Women
Ammunition (National)	5.4	15.8
Explosives (National)	3.8	15.3
Instruments and Optical Equipment (National)	6.3	36.6
All Reporting Manufacturing Establishments (National)	6.2	22.3
Manufacturing Establishments in Elgin, Illinois	5.2	20.7
Company Plant No. 2 in Elgin, Illinois	3.42	65.0

* Sources: National Figures: Bureau of Labor Statistics, U. S. Dept. of Labor.
Elgin Figures: Personal Survey by the writer.

These figures are of special interest mainly because of the fact that in Company Plant No. 2, although the percentage of women is much higher than in the other industries, the absence rate is substantially lower than the average. In other words, the total situation is very good when compared to both area and national figures.

Summary and Conclusions

In general, as a result of the study, the following facts become evident:

1. Women have three times as much absenteeism as men, in total rates.
2. Women have approximately twice as much sickness absenteeism as do men.

3. Women have between three and four times as much personal absenteeism as do men.
4. It is apparent that these differences are the result of the sex variable, since they are evident in every age and service group and department.
5. In general, sickness and personal absences were nearly evenly divided.
6. Age groups show no great difference in rates, although there is a slight tendency for the older employees to be absent more.
7. Service groups show more variation, with the rates tending to increase as service increases, up to a point at about six months of service.
8. Final Assembly Department has the highest rates in both sickness and total absences of all the major departments.
9. A striking difference is shown in the case of the personal absences for women in the Plate Department, the rate here being much higher than in any other major department.

In conclusion, it would seem that the findings of this study on the subject of sex differentials tend to agree with other similar investigations as reported in the literature. The age groups show no tendency toward being of importance in influencing the absence rates, and the length of company service factor would appear to be of more importance here from the standpoint of influence upon absenteeism.

Probably the most important findings of the study aside from these mentioned above, are in the uncovering of the two striking departmental differences, that is, Final Assembly in total absenteeism, and Plate in personal absences for women. While the results in the case of Final Assembly are not such that it is possible to give the answer to this situation, they do serve to focus attention upon the problem and to disprove one possible hypothesis; namely, that the younger workers are causing the majority of the absences in this department.

Received November 3, 1943.

Bibliography

1. American Business, March, 1943. "Find Chronic Offenders to Cut Absenteeism." P. 13.
2. ——, April, 1943. "Our Errors in Fighting Absenteeism." P. 13.
3. British Information Services, special pamphlet, April, 1943. "Working Conditions and Absenteeism in Britain." 10 pages.
4. British Ministry of Labour and National Service, special pamphlet, 1942. "Problem of Absenteeism," 8 pages.
5. Factory Management and Maintenance, special reprint pamphlet, July, 1943. "Tested Ways to Reduce Absenteeism." 25 pages.
6. Management Review, November, 1940. "Cuts Lost Time Due to Colds." P. 300.
7. ——, September, 1942. "How War Industries Fight Absenteeism." Pp. 311–312.
8. National Industrial Conference Board, Inc., *Studies in Personnel Policy*, 23, New York, 1940. "Personnel Practices in Factory and Office." 55 pages.

9. ———, Studies in Personnel Policy, 46, New York, 1942. "Reducing Absenteeism." 35 pages.
10. Spriegel, W. R., and Schulz, E. *Elements of supervision*, John Wiley and Sons, Inc., New York, 1942. Pp. 150–167.
11. United States Department of Labor, Bureau of Labor Statistics; *Monthly Labor Review*, January, 1943—Vol. 56, No. 1. "Problems of Absenteeism in War Production." Pp. 1–9.
12. ———, *Monthly Labor Review*, July, 1943, Vol. 57, No. 1.
13. ———, Special Bulletin No. 12-A. "Auditing Absenteeism." 30 pages.
14. War Production Drive Headquarters, War Production Board, Washington, D. C., special pamphlet. "Absenteeism Guide for Plant Labor-Management Production Committees." 27 pages.
15. Watkins, Gordon S., and Dodd, P. A. *The management of labor relations*, McGraw-Hill, New York, 1938. Pp. 250–275.

Testing the Pulling Power of Advertisements by the Split-Run Copy Method *

Joseph Zubin

Columbia University

and

John G. Peatman

The City College of New York

Many different methods are employed by the advertising industry for the purpose of determining the effect of advertising copy on sales. One of the most common methods is that of "split-run copy" testing.

The method itself can be briefly described as one in which two (or more) forms of a given advertisement are printed in a newspaper or magazine of a given issue, the advertisements being alternated in the production of the publication medium so that the different forms will be randomly distributed to the reading public. When properly carried out, such a method should achieve the desired end of securing two randomly selected groups of the population under study, equal in size, one of which is exposed to the first form of the advertisement and the other to the second form.

The relative pulling power of each form of the copy is then measured by the number of replies received. All copy tests therefore need to include an offer of some article such as a free sample of the product or a souvenir. This article should be sufficiently attractive to call forth a volume of response large enough to be subjected to a statistical test of the significance of the difference in pulling power between the two forms of the advertisement. The free offer needs to be towards the end of the advertising copy and relatively inconspicuous in order to insure that the response is brought about by the reading of the copy itself rather than by the attractiveness of the free sample or souvenir alone.

An example of a split-run copy test is given in an article by Manville (3) in which the problem was "to determine the relative pulling power of the words 'False Teeth' versus the words 'Dental Plates' in headlines" of copy advertising Polident, a cleanser for false teeth. Two split-run tests were made, one in the New York Times Sunday Magazine Section

---

* The authors wish to thank Miss Jane E. Farwell for drawing the nomographs included in this paper.

and the second in the New York News Sunday Rotogravure Section. The results obtained are summarized by the author as follows:

1. *The New York Times split-run copy test, made March 9, 1941,* yielded a number of replies, 51.4 per cent of which were from the F.T. ("False Teeth") copy and 48.6 per cent were from the D.P. ("Dental Plate") copy. The author states that "this is what may be called an inconclusive result; where separation between the competing advertisement shows less than ten points (not 10 per cent) difference.[1] Polident's experience has shown that no material difference exists between two tested advertisements. However, in conclusion, note here that 'False Teeth' was a shade better."

2. The results of the second split-run test made in the N. Y. News Sunday Rotogravure Section on September 20, 1942, yielded a set of replies, 52.5 per cent of which were from the F.T. copy and 47.5 per cent from the D.P. copy. The author presents these results, stating "here is another test run to verify results from the first test. . . . Again 'False Teeth' was a shade stronger—almost mathematically perfect."

Since the author gives only the percentage of returns for each copy but does not give the total number of returns, it is impossible to evaluate the significance of his results on a statistical basis. Furthermore, even if the author had given the absolute frequencies instead of the percentages we would still be unable to treat the results statistically since we have no knowledge of the number of readers who are potential buyers of the product but failed to respond. The latter data are rarely if ever known even to the research worker. If we wish to evaluate the above results despite this deficiency, certain assumptions must be made regarding the number of replies as well as the total number of readers who are potential buyers of the product advertised.

We shall present in this paper several general methods for evaluating the results and apply them to Manville's data as an example. In order to test the significance of his results, we shall proceed on two alternative assumptions:

Situation (A): that he received the minimum average cited by Sturgis (4) of 100 replies per tested advertisement or 200 for both;

Situation (B): that he even may have received as many as 500 replies, on the average, for each of his tested advertisements or 1000 for both.

Inasmuch as the greatest difference in the two split-run tests made by Manville was obtained from the Daily News, we shall confine our examples to those figures, namely, 52.5 per cent for the F.T. copy and 47.5 per cent for the D.P. copy.

---

[1] It is clear that on statistical grounds alone, this statement is not tenable. The absolute difference between two per cents can not be evaluated directly, but must be referred to its standard error.

If Manville received a total of 200 replies in this test, he would have received 105 (52.5%) from the F.T. copy and 95 (47.5%) from the D.P. copy.

If, on the other hand, we assume that he received as many as 1000 replies, he would have received 525 from the F.T. copy and 475 from the D.P. copy.

With an estimate of the actual *number* of replies received for each copy, we are in a position to test the significance of the results, that is to say, whether the F.T. copy really had more pulling power than the D.P. copy or whether the difference is such as to be attributable to chance. Unless we assume, however, that the replies were obtained from a very large sample [2] of potential buyers of the product exposed to the advertisements, we will need to estimate the size of the sample, N, before we can make a statistical test for the significance of the result. We shall illustrate the development of the test of significance for both circumstances, that is, the one in which we assume a very large sample (100,000 or more) and the other where the sample may be relatively small (say 2000).

Both procedures assume that the two versions of the copy to be tested are equally and randomly distributed among readers who are potential buyers of the product. It should be apparent that an estimate of the size of such groups would not in this particular case be equal to the total circulation of the paper on the day of the test. Not every one of the readers of the paper would be a potential buyer of the cleanser. The actual circulation of the News on the day of the test was reported as 2,175,429. It is, therefore, safe to assume that the actual number of potential buyers exposed to the copy was considerable.

Table 1

Hypothetical Distribution of Responses to the F.T. and D.P. Copies

Potential Buyers Exposed to Copy	F.T. Copy	D.P. Copy	Totals
Responded	a	c	$a + c$
Did not respond	b	d	$b + d$
Totals	$N/2$	$N/2$	N

Assuming then that the two versions of the advertisement reached groups which were equally saturated with potential buyers, we can draw up the hypothetical two-by-two table shown in Table 1.

---

[2] Large as compared to the number of returns. The discussion of the importance of the relative size of N will be treated later. Note that the "Sample" includes all potential buyers of the product who were exposed to the advertising copy, and not merely the number of those actually replying.

where a is the number of potential buyers who read the F.T. copy and *responded*, b the number of potential buyers who read the same copy and *did not respond*. The letters c and d represent the corresponding data for the D.P. copy; N is the total number of readers who are potential buyers, half of whom, $N/2$, read the F.T. copy and half the D.P. copy.

Let us define the "pulling power" of an advertisement as the proportion of potential buyers who read the copy and were sufficiently moved to mail in the coupon at the end of the copy. The pulling power of the F.T. copy is the proportion $a/(N/2)$, and for the D.P. copy, $c/(N/2)$, or $2a/N$ and $2c/N$ respectively.

In order to determine whether the difference between the two "pulling powers" is significant, we apply the simple test for the significance of the difference between two per cents. The critical ratio of this difference, CR, is:

$$\text{(I)} \qquad CR = \frac{2a/N - 2c/N}{\sqrt{pq[2/N + 2/N]}} = \frac{(a - c)}{\sqrt{pqN}}$$

where $p = (a + c)/N$ and $q = 1 - (a + c)/N$.

This equation is somewhat different from the usual one given in some elementary texts but it is the more correct form and the justification of its use is given elsewhere (5).

$$\text{(II)} \qquad \text{Hence, } CR = \frac{a - c}{\sqrt{(a + c)[1 - (a + c)/N]}}$$

In order to remove the square root sign, we can square both sides of the equation and obtain the expression for Chi-square.

$$\text{(III)} \qquad (CR)^2 = \chi^2 = (a - c)^2 \Big/ \left[(a + c)\left(1 - \frac{a + c}{N}\right)\right]$$

Fisher and Yates (2) have pointed out that whenever the smallest expected frequency (number of responses expected for a given copy when chance alone or some other definite hypothesis is assumed to be operative) is less than 500, a correction for continuity (designated by them as the Yates correction) should be applied. This consists of simply reducing the net value of $(a - c)$, the difference between the number of responses for the two copies, by unity. Hence equation (III) becomes

$$\text{(III}'\text{)} \qquad \chi^2 = [|a - c| - 1]^2 \Big/ \left[(a + c)\left(1 - \frac{a + c}{N}\right)\right]$$

Let us now apply this equation to test the significance of Situation A where the total responses were 200, 52.5 per cent for the F.T. copy and 47.5 per cent for D.P. copy. First, however, we must make some assumption regarding the sample size, N, the total number of potential buyers who read the copy, regardless of whether they responded or did not respond.

Case I—Few Replies from a Large Sample: Number of returns very small compared to the total number of potential buyers exposed to copy.

Let us assume for Case I that the number of returns is negligible (less than 1%) compared to the total number of readers who are potential buyers. In this instance, equation (III′) reduces to a much simpler form, as follows:

If the proportion, $(a + c)/N$, is negligible, or approaches zero, then

$$\text{(IV)} \qquad \chi^2 = (|a - c| - 1)^2/(a + c)$$

and for situation A: $\chi^2 = (9)^2/200 = 0.40$, $P = .50$,
and for situation B: $\chi^2 = (49)^2/1000 = 2.40$, $P = .12$.

In both of these situations the difference is not statistically significant. Hence, under the above assumption of a very large sample of readers who are potential buyers, the hypothesis that the two advertisements were *equal* in pulling power is quite tenable; consequently, Manville's implication that the F.T. copy was really more effective would have to be rejected.

We might reverse the question and ask how large should the difference have been in order to produce a significant difference for situations A and B. Accepting a value of $\chi^2 = 6.635$ ($P = .01$) as the lower limit of significance, we must solve for a and c, with $(a + c)$ equal to 200 in situation A and to 1000 in situation B. Hence, we solve equation IV for the value $(|a - c| - 1)$, as follows:

$$\text{(V)} \qquad (|a - c| - 1)^2 = \chi^2(a + c)$$

After determining the value of $a - c$, we can readily determine a and c respectively, since $(a + c)$ is known.

For situation A: $a = 119$ and $c = 81$ (or $a = 59.5\%$ and $c = 40.5\%$).
For situation B: $a = 541$ and $c = 459$ (or $a = 54.1\%$ and $c = 45.9\%$).

Consequently for situation A, if one copy had brought 59.5% or more of the responses (and the second 40.5% or less), the difference in the respective pulling powers would have been significant.

Similarly in situation B, if one copy had pulled 54.1% or more of the responses (and the second 45.9% or less) the difference would have been significant.

We have prepared a nomograph based on the relationship of equation (V).[8] It has been plotted for values of total volume of responses from 100 to 10,000 and for *differences* in responses between two copies, from 11 to 100.

---

[8] We decided to deal with absolute frequencies in equation V rather than with per cents (which are more generally used by advertising men) because the equation in its per cent form is much more complicated than in its absolute frequency form. The

Figure 1 is read as follows: When a total number of potential buyers exposed to the copies is extremely large, then for any given volume of responses $(a + c)$ there will be a range of possible differences in response $(a - c)$ to the two copies, some of which will be significant. To determine whether the difference between a given set of replies is significant, proceed as follows:

The total volume of responses $(a + c)$ is found on the line to the right, and the difference between the frequency of response for the two copies is found on the line to the left. By joining these two points with a ruler the exact value of P can be read from the middle line. This value of P indicates how often a difference as large as (or larger than) the one observed could arise by chance. When this difference is so large that it can arise by chance less than 1 time in 100 $(P = .01)$ the difference is regarded as statistically significant. When the value of P or the observed difference lies between .05 and .01, the result is doubtful, and when P exceeds .05, the difference is regarded as insignificant. For example, when the total volume of responses is 200 and the difference in replies between the two copies is 19, the value of P found from the nomograph is about .22 and hence not significant.

The nomograph of Figure 1 can also be used to determine how large the total volume of responses must be in order that a given difference in replies will be significant. Thus, when the difference is 50, there must be *no more* than 375 total responses if the difference is to remain significant, since the significance of a given absolute difference *decreases* as the number of replies *increases*.

The nomograph of Figure 1 may also be used to determine the minimal size of a difference in replies between two copies that will yield a significant difference for a given volume of responses. Thus, for a volume of 500 responses there must be at least a difference of 58 in the two sets of replies, for the result to be significant.

---

former may be written as follows:

(V′) $$\chi^2 = [|p_1 - p_2| - 1/(a + c)]^2(a + c)$$

In order to take care of the factor $1/(a + c)$ a double nomograph would have to be used instead of the single nomograph which is sufficient for equation (V). Had we neglected the $1/(a + c)$ factor completely we would have overestimated the value of χ (the root of χ^2 which equals the value of the critical ratio) by as much as $1/\sqrt{a + c}$. For 100 replies, χ is overestimated by .1. This is a small enough error, but its size was not realized until after the nomographs were drawn. The nomograph in its per cent form has since been drawn up and may be obtained from the authors.

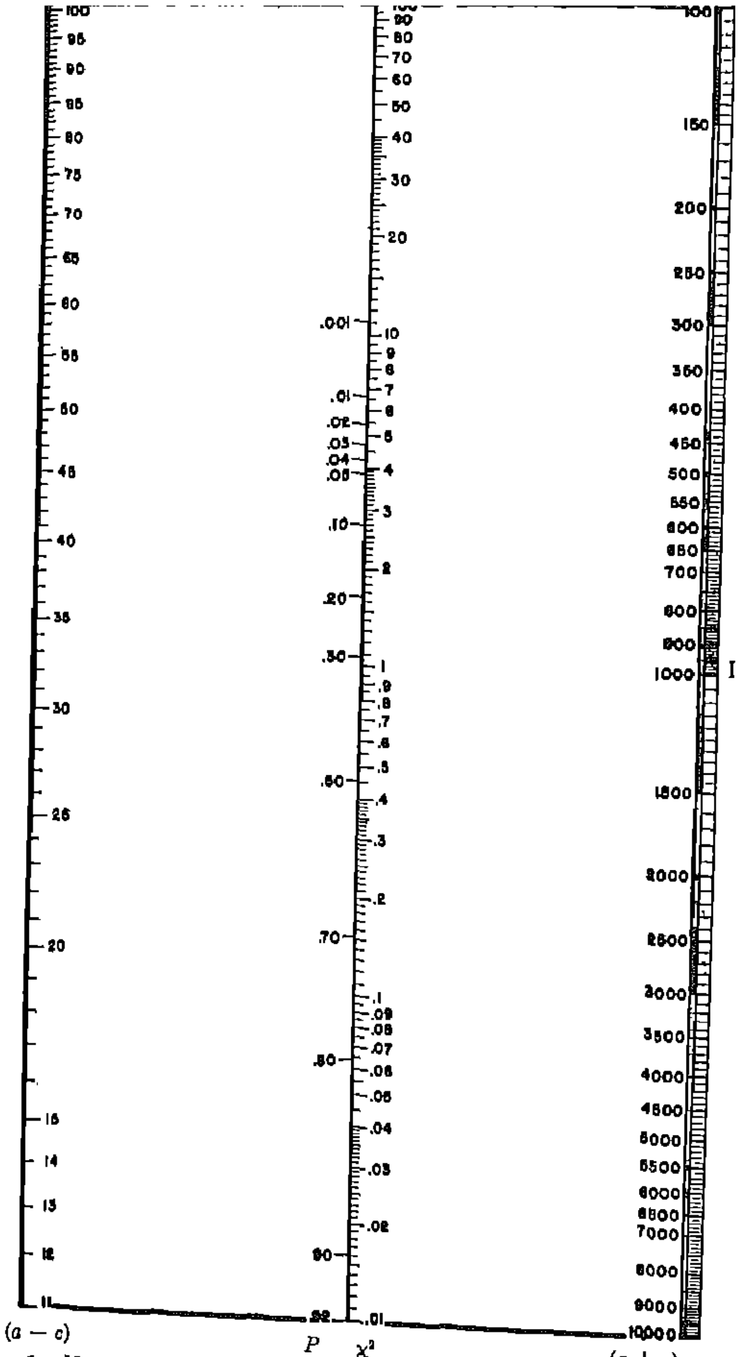


FIG. 1. Nomograph I for determining the significance of the difference in number of responses to two contrasted copies when the total number of responses is less than 1% of total number of potential buyer-readers.

We can also determine the significance of the results obtained from combining two independent split-run copy tests. Thus, if we were to combine the values of χ^2 given by Nomograph I for the results yielded by the New York *Times* and the *News* we would obtain the following results for situation A and situation B (1).

Table 2

Significance of the Difference in Response to the F.T. and D.P. Copies for the Combined Results of the Split-Run Copy Tests in the Times and the News

	For Situation A $(a + c) = 200$* χ^2	D.F.	*For Situation B* $(a + c) = 1000$ χ^2	D.F.
Times	.125	1	0.74	1
News	.405	1	2.40	1
Total χ^2	.525	2	3.14	2
P	.80		.20	

* Had to be computed directly because nomograph did not extend to such low values. D.F. represents degrees of freedom.

Under the assumption that a total of only 200 responses was received for the F.T. and D.P. copies (situation A), the combined results for the difference in response in both newspapers would arise by chance about 80 times in 100 and such a difference is not statistically significant. Under the assumption that 1000 responses were received (situation B), the combined results indicate that chance alone would account for the difference 20 times in 100, which is again not significant.

It should be pointed out here that since the total number of responses is taken to be at least 200 for situation A, the number of responses expected by chance for each copy is 100. Fisher and Yates (2) have indicated that when the smallest expected frequency is not less than 200, the value of χ^2 when corrected for continuity will give a good approximation of the true probability. However, when the number of total responses falls much below 200 the method of χ^2 usually fails to give a good approximation to the true probability. In the case under discussion, however, since the two contrasted groups of readers of both copies are considered equal in number, the stringency of the above rule can be relaxed. It has been found in practice that when the two contrasted groups are equal, the method of Chi-square is still applicable even when the smallest expected frequency is as low as 5, that is, when the total number of responses is only 10.[4]

---

[4] The reader may wonder whether it is fair to consider only the two columns of our 2 × 2 table in judging whether the contrasted groups are equal. When we compare the two rows, the contrasted groups are far from equal. However, the value of χ^2 is independent of whether we compare the rows or the columns. Hence, equality in either is sufficient.

Case II—Replies from Small Samples: Number of returns large compared to the total number of potential buyers exposed to the copy.

We can now turn our attention to the second possibility previously suggested according to which the number of returns represents a significant portion of the total number of potential buyers exposed to the copy. Let us accept a sample of 2000 as such a number, 1000 for the F.T. copy and 1000 for the D.P. copy. Then the 200 responses of situation A would constitute 10% of the total sample, and the corresponding proportion for the 1000 replies of situation B would be 50%. We can now construct Table 3 for situation A, Case II.

Table 3

Data for Situation A, Case II (200 replies from a sample of 2000)

Potential Buyers Exposed to Copy	F.T. Copy	D.P. Copy	Totals
Responded	105	95	200
Did not respond	895	905	1800
Totals	1000	1000	2000

Table 3 is constructed as follows: First, Manville's values of 52.5% (or 105 returns for the F.T. copy) and 47.5% (or 95 returns for the D.P. copy) are entered in the first row of the table together with the total volume of 200 replies. Since the total sample of potential buyers is taken as equal to 1000 for each copy, the differences between 1000 and the number of returns for each copy are entered in the second row of the table. Finally, the marginal totals are entered.

For Table 3, $\chi^2 = .45$ with a P value of .50, and hence the difference in pulling power between the two copies is not significant.

For situation B, Case II, we obtain the results of Table 4.

Table 4

Data for Situation B, Case II (1000 replies from a sample of 2000)

Potential Buyers Exposed to Copy	F.T. Copy	D.P. Copy	Totals
Responded	525	475	1000
Did not respond	475	525	1000
Totals	1000	1000	2000

Here again the difference falls short of being significant, $\chi^2 = 4.8$ ($P = .29$).

In other words, the difference in the pulling power of the two copies is again found to be non-significant for both situations A and B, even when the total sample of exposures is small.

We can now reverse the problem and determine when a difference in pulling power will be significant if the total sample consists of only 2000 potential buyers exposed to the copy, 1000 for the F.T. copy and 1000 for the D.P. copy, and with the total number of returns taken as 200 for situation A and 1000 for situation B.

Solving equation (III′) for $(|a - c| - 1)^2$, we obtain

$$\text{(VI)} \qquad (|a - c| - 1)^2 = \chi^2(a + c)\left[1 - \frac{a + c}{N}\right]$$

Since we know the value of $a + c$, we can solve the equation readily. For situation A (200 replies from a total sample of 2000): $a - c = 35.56$; and $a + c = 200$; $a = 118$; $c = 82$. Hence if the total sample of exposures were only 2000 and the responses to the F.T. copy were 118 or more (and for the D.P. copy 82 or less) the difference would have been significant.

For situation B (1000 replies from a total sample of 2000): Solving equation (VI): $a = 530$ and $c = 470$. Hence, if 530 responses or more had been received to the F.T. copy (and 470 or less for the D.P. copy) from a total sample of only 2000, the difference would have been significant.

Case III—Results for a Sample of 4000 Exposures.

Let us now consider situation A and B, but instead of a total sample of 2000 potential buyers exposed to the copy let us assume a total of 4000 such exposures. Paradoxically, the value of χ^2 drops from .45 when $N/2 = 1000$ to .37 when $N/2 = 2000$ for the 200 replies of situation A, and from 4.8 to 3.20 for the 1000 replies of situation B. Both of these values of χ^2 are below the critical value of significance ($\chi^2 = 6.635$), and hence the difference cannot be considered significant. But in both cases the value of χ^2 decreased as N increased. This is perhaps unexpected since usually the value of χ^2 increases as N increases or, in other words, the significance of a given difference increases as the sample size increases. We must remember, however, that our problem is somewhat unusual, since for a given number of replies the pulling powers, $2a/N$ and $2c/N$, decline as N, the size of the sample, increases and consequently the numerator in equation (III) also declines. On the other hand, the denominator increases as N increases since the expression $1 - (a + c)/N$ will increase as N increases, providing $(a + c)$ remains unchanged. Thus, with the numerator decreasing and the denominator increasing, the value of the ratio, χ^2, must perforce decline.

However, the value of χ^2 cannot decline indefinitely, for as N increases indefinitely, the value of χ^2 approaches $(|a - c| - 1)^2/(a + c)$ as a limit. Therefore, if the limiting value of χ^2 is greater than the value 6.635, increases in the value of N will not reduce the significance of the difference, but decreases in N will increase the significance of the result. On the other hand, if the limiting value of χ^2 is less than 6.635, a decrease in N of sufficient size may produce a significant difference.

Case IV—The Maximum Size of Sample that will yield a Significant Difference for a Given Number of Replies.

Since we have determined that the value of χ^2 decreases for a given number of replies as N, the size of the total sample, increases, we may now reverse the problem and inquire how small N must be in order to produce a significant difference between the pulling power of two advertisements in a split-run test.

(VII) Solving equation (III) for N, we obtain

$$N = \frac{(a+c)^2}{(a+c) - \dfrac{(|a-c|-1)^2}{\chi^2}}$$

or

(VII′)
$$1/N = 1/(a+c) - \frac{(|a-c|-1)^2}{(a+c)^2\chi^2}$$

Applying this formula we find that N is 213 for A. Hence when the total volume of replies is approximately 214, or 107 for each copy, the results are significant in favor of the F.T. copy. When the volume increases above 214 the results are no longer statistically significant. But what happens when the sample falls below 214? Obviously it cannot fall indefinitely below 214 since the total sample can never be less than the total volume of responses. In fact, the minimum value for the sample size is twice the number of responses obtained from the copy with the larger number of responses. This limiting situation is the one which produces maximum differentiation between the two copies. This optimum circumstance is obtained when all the readers of one copy respond (a hypothetical situation which will hardly ever occur). In the case of situation A, the optimum circumstance for a possible significant difference is shown in Table 5.

Table 5

Optimum Circumstance for a Possible Significant Difference in Situation A (200 replies)

	F.T. Copy	D.P. Copy	Total
Responded	105	95	200
Did not respond	0	10	10
Total	105	105	210

In Table 5 there is a total of 210 readers of whom half, 105, read the F.T. copy and half the D.P. copy. Of the 105 who read the F.T. copy, *all* responded; while of those who read the D.P. copy, 10 failed to respond. The value of Chi square for the above table is 8.51. Thus, when the volume of responses is 200, of which 105 come in response to the F.T. copy and 95 to the D.P. copy, we may conclude that the F.T. copy is superior in pulling power to the D.P. copy as long as the total sample size lies between 210, the minimal possible size, and 213. When the total sample goes much beyond 213, the difference is no longer statistically significant.

For situation B (1000 replies) the optimum result is shown in Table 6.

Table 6

Optimum Circumstance for a Possible Significant Difference in Situation B (1000 replies)

	F.T. Copy	D.P. Copy	Total
Responded	525	475	1000
Did not respond	0	50	50
Total	525	525	1050

$\chi^2 = 50.42$ and hence the difference in the pulling power of the F.T. and D.P. copy of Manville's Daily News split-run test would have been statistically significant if the total sample of potential buyers exposed to the copy were only 1050 and the total number of replies received were 1000. Such a large number of replies as 1000 may very well be unlikely for the one insertion of the advertisements. In any event, it is even more unlikely that the more than two million circulation of the Daily News included not more than 1050 readers who were potential buyers of Polident.

We can now ask how large may N grow above 1050 before the difference becomes non-significant. Substituting in equation (VII),

$$N = \frac{(1000)^2}{1000 - \frac{(49)^2}{6.635}} = 1566$$

Hence, when N is 1566, or 783 readers for each copy, the difference between 525 responses for the F.T. copy and 475 for the D.P. copy will still be significant. However, when the sample exceeds 1566, the difference in pulling power of the two copies becomes statistically doubtful. In other words, as long as the total number of readers (evenly divided between the two copies) lies between 1050 and 1566, the difference in pulling power for the 1000 replies of situation B is significant.

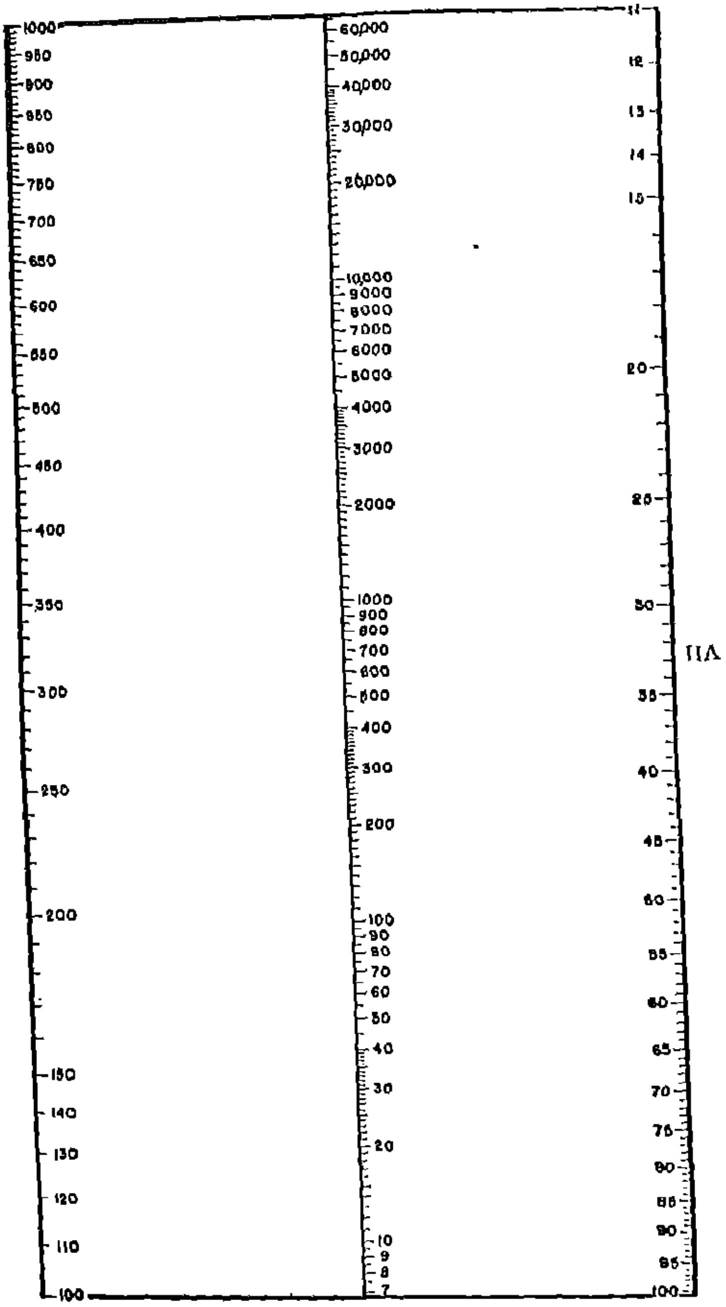
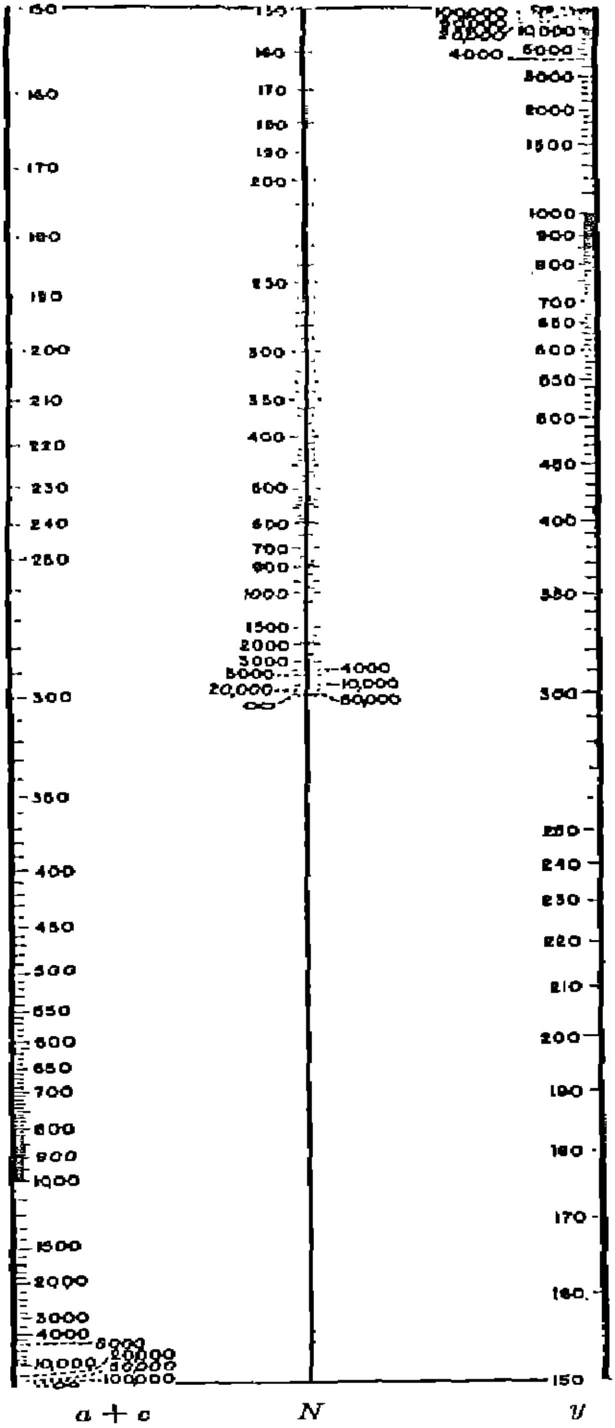


FIG. 2. Nomograph IIA for determining the number of potential buyer-readers (N) required to produce a significant difference for a given total frequency of response, $(a + c)$, and a given difference in frequency for two contrasted copies, $(a - c)$.


a + c
N
y
150
160
170
180
190
200
210
220
230
240
250
300
350
400
450
500
550
600
650
700
800
900
1000
1500
2000
3000
4000
5000
10,000
20000
50000
100000
20,000
∞

IIB

When N increases to about 1500, the difference begins to fall below the significance level. Thus, when we have some estimate of the number of readers exposed to split-run copy, we can determine for a given set of replies whether or not the difference in pulling power is significant.

Two nomographs (IIA and IIB) have been developed which will indicate the maximum value of N that will yield a significant difference between observed results. These are presented in Figures 2 and 3. They are based on equation (VII′). Letting $y = (a + c)^2\chi^2/(|a - c| - 1)^2$, equation (VII′) becomes:

$$\text{(VII″)} \qquad 1/N = 1/(a + c) - 1/y$$

We can draw up a nomograph for determining the value of y from the values of $(a + c)$ and $(a - c)$ for the situation when $\chi^2 = 6.635$ ($P = .01$). This nomograph (IIA) is shown in Figure 2.

The left hand column gives the total number of responses $(a + c)$ and the right hand column gives the difference in responses between the two copies $(a - c)$. The middle column gives the value y which is to be used in Nomograph IIB. A ruler connecting the two points for $(a + c)$ and $(a - c)$ respectively will cut the middle line at the desired value of y. Having determined the value of y, we enter with this value of y on the line to the right in the Nomograph IIB and by connecting the point y on that line by means of a ruler with the point for the total number of responses $(a + c)$ on the line on the left, we can read off in the center the value of N required to render the difference significant.

Example for Nomographs IIA and IIB: When the total number of responses to both copies is 1000 and the difference in responses between the two copies is 50, we enter Nomograph IIA in the left hand column marked $(a + c)$ with the value of 1000 and connect this point with the ruler to point 50 on the right hand scale and read off the value of y on the middle scale as 2650. With this value of y we enter the right hand scale of Nomograph IIB and connecting it up with the point corresponding to 1000 on the left scale, we read off the value of $N = 1500$ on the middle scale. That is to say, when the total number of responses is 1000 and the difference in response between the two copies is 50, a total sample of not more than 1500 potential buyer-readers would be required to make the difference significant.

Nomographs IIA and IIB can be used *in any one of four ways*. First, to determine the maximum value of N that will yield a significant difference for a given set of replies, as was done in the example above. Second, to determine that value of a difference in replies $(a - c)$ which would be significant for a given size sample (N), and a given volume of replies $(a + c)$. Third, to determine the volume of responses $(a + c)$

which would yield a significant difference for a given difference in replies $(a - c)$ and for a given sample size (N). Fourth, to determine the significance of the observed difference by obtaining the value of χ^2 directly. In order to utilize the nomograph for this purpose, only one computation is required.[6] First, determine from Nomograph IIB the value of y_B corresponding to the observed values of $(a + c)$, the total number of replies and (N) the total sample. Secondly, determine the value of y from Nomograph IIA which gives the value of y_A corresponding *to the observed number of replies $(a + c)$ and the difference in replies* $(a - c)$. The desired value of χ^2 is obtained from the following formula:

$$\chi^2 = \frac{y_A}{y_B}(6.635)$$

Knowing the value of χ^2, we can determine P from any table of χ^2.

Discussion

Thus far we have dealt only with the statistical aspects of the split-run copy technique. However, it is to be emphasized that there are certain assumptions which the technique must satisfy before the statistical method becomes applicable. These assumptions are:

(1) That the two copies of the advertisement have been so distributed among two groups of readers that these groups constitute random samples of the population under examination—randomly divided at least insofar as the product being advertised is concerned.

(2) That the two groups contain an approximately equal number of potential buyers of the product. (The method can also be applied when the potential number of buyers of the two groups is not equal; in such case, however, their relative number must be known.)

(3) That a valid estimate of the maximum size of the sample of potential buyers is available. It should be borne in mind that no amount of statistical refinement can compensate for any gross error in the original estimate of the size of the samples under comparison.

---

[6] A more direct way of accomplishing this purpose is to note the fact that Nomograph IIA is based on the value of $\chi^2 = 6.635$ as the critical point. This value of χ^2 is found on the y scale opposite the value $31.6 = (\sqrt{1000})$ on the $(a + c)$ scale. (The decimal point has to be provided by the reader.) By providing a duplicate y scale and moving up or down we can let the critical value of χ^2 vary from 6.635 to any other value we please. After determining the value of y for a given observed result by means of Nomograph IIB, we can enter Nomograph IIA and let the duplicate scale move up or down until the value of y_B lies under the ruler connecting up $(a + c)$ with $(a - c)$. The value of χ^2 for the observed comparison can then be read off at the critical point *(corresponding to 31.6 on the $(a + c)$ scale)*, with *proper* care for the decimal point. The corresponding value of P can be obtained from standard tables.

(4) That the act of clipping the coupon at the end of the copy is indicative of interest in the product, and furthermore, that this act was brought about as a result of the reading of the copy, and not as a result of any other factor.

When two contrasted copies are found to yield results which are significantly different, care must be taken not to place too much reliance on this fact alone. The statistical procedures described serve to indicate whether a difference in a given set of replies is or is not greater than would be expected on the basis of chance alone. The attributing of a superiority in pulling power to one copy over the other must be buttressed by evidence of a logical or a psychological kind. If the result flaunts common sense, the test should be repeated until it is verified beyond any doubt.

If it stands up under repetition, it would be well to investigate the causes of the differences before the superiority of one form over the other is regarded as established. Simple methods for combining the results of two or more tests are presented by Fisher (1) and an example has been worked out in this paper.

Application to Other Problems

The nomographs described in this paper are useful in other situations besides the split-run copy technique. They will be found useful in item analysis of psychological tests when the two contrasted groups of successes and failures are equal. For this purpose, Nomographs IIA and IIB should be used. The advantage of these nomographs over previous ones inheres in the fact that the significance of an item can be read off directly from the nomographs once the values of $(a - c)$, $(a + c)$ and N are known.

Whenever the incidence of rare events is contrasted in two groups, Nomograph I can be used to determine the significance of the difference in the incidence of the event, if the two contrasted groups are very large and equal in size. Examples of studies in which many rare events or characteristics are compared in two contrasted groups are quite plentiful. Among these are studies in the incidence of rare diseases and accidents in different racial or industrial groups. Comparison of the incidence of rare words in contrasted authors or manuscripts is another example. Nomograph I offers only an approximation to the value of the exact probability and should be useful as a screening device for selecting the comparisons that need further study with the exact methods developed by Fisher (1) or by means of the Poisson distribution which is suitable for the evaluation of the statistics of rare events.

Summary

In determining the relative pulling power of two different forms of the same advertising copy, a method known as the split-copy testing has been widely used. Heretofore only rule of thumb methods have evidently been widely used in the evaluation of the results of this technique. The present article develops simple formulae which are modifications of the basic Chi-square formula for equal contrasted groups (marginal frequencies in the columns, or in the rows). One of the difficulties characteristic of many studies in this field is the absence of any precise data regarding the actual size of the number of readers of the advertisements in question who are potential buyers of the article advertised. Consequently it becomes necessary to indicate maximal and minimal sizes of samples which will yield statistically significant results for a given set of returns. Three nomographs are presented for eliminating all computations and obtaining the required answers directly. These nomographs can be used for any situation in which a comparison is made between the frequencies of an event (a reply, in the case of advertising copy tests) for two equal contrasted groups. Nomograph I is to be used when the frequency of the event under investigation is small (less than 1%) as compared to the total sample. Nomographs IIA and IIB are to be used when the frequency of the event is considerable (higher than 1% of the total sample). The use of these nomographs has the advantage of requiring no computations whatsoever, the significance of the results being determined directly from the raw data.

An example from the literature of a duplicated split-run test for comparing the pulling power of two advertisements was carefully analyzed. The results were found to be statistically non-significant for each test separately as well as for both tests taken in combination.

Received January 3, 1944.

References

1. Fisher, R. A. *Statistical methods for research workers.* London: Oliver and Boyd, 1934.
2. Fisher, R. A., and Yates, F. *Statistical tables for biological, agricultural and medical research.* London: Oliver and Boyd, 1938.
3. Manville, R. How to test copy: An example in its simplest form. *Printers Ink*, 1943, 202, 17–19 and 87.
4. Sturgis, W. A. Blue print of copy testing. *Printers Ink*, 1943, 203, 20–23.
5. Zubin, J. Note on a graphic method for determining the significance of the difference between group frequencies. *J. educ. Psychol.*, 1936, 27, 431–444.

For other nomographs applicable to the 2 × 2 or 2 × n table see:

6. Zubin, J. Nomographs for determining the significance of the differences between the frequencies of events in two contrasted series or groups. *J. Amer. Statis. Asso.*, 1939, 34, 539–544.
7. Fulcher, J. S., and Zubin, J. The item analyzer: A mechanical device for treating the four-fold table in large samples. *J. appl. Psychol.*, 1942, 26, 511–522.

Relationships Between Strong Vocational Interest Scores and Other Attitude and Personality Factors

Leona E. Tyler

University of Oregon

At the present stage in the progress of vocational interest research, the most challenging problems are those concerned with the nature of the characteristics which differentiate one occupational group from another. A vast mass of information has accumulated with regard to empirical differences between groups and relationships between different interest scores (4). So far only a few studies have attempted to clarify and explain these relationships or to fit the traits under consideration into a general theory of personality.

Purposes of This Study

The chief aim of the present study was to analyze in some detail relationships between scores on the *Strong Vocational Interest Blank* and several measured attitude and personality factors. This involved going behind the correlations to discover, if possible, *why* the scores correlated. It was hoped that such a procedure would throw light on the nature of differential occupational interests and would also add to our understanding of personality organization. It is to be remembered that the interest tests are practically the only so-called personality tests which have any objective outside reference. We know that interest scores reflect choices people actually make and plans of action they actually carry out. To be able to relate other measured characteristics about which we know less to this reference point would have some real advantages. For the author, this study is related also to plans for a longitudinal investigation of the development of interests in children. Exploration of the relationships obtaining for adults shows what characteristics are important to observe in a genetic study.

Procedure

The subjects were college sophomores taking psychology laboratory at the University of Oregon in the fall term of 1941–1942, 55 men and 122 women. The battery of tests consisted of the *Strong Vocational Interest Blank*, the *Minnesota Personality Scale* with its five sub-tests measuring *Morale, Social Adjustment, Family Adjustment, Emotional*

Adjustment, and *Economic Conservatism*, and the Thurstone and Chave scales, *Attitude toward the Church* and *Attitude toward God* (No. 22, Form A). The Strong tests were scored on the following group scales: Group I. Human Sciences; Group II. Technical Sciences; Group V. Personnel or Social Service; Group VIII. Office Work; Group IX. Sales; and Group X. Verbal-Linguistic.

The first step in the analysis was the computation of correlations between each of the interest scores and each personality and attitude variable, keeping the data for men and women separate. These coefficients were checked for statistical significance by Fisher's "*t*." All relationships for which the probability of significance was more than .95 were selected for further analysis. This consisted of an item analysis of the personality measurement under consideration. We singled out the individuals receiving highest and lowest scores on the *interest* scale. In the men's group, we compared the 20 highest with the 20 lowest. In the women's group we used the 33 highest and the 33 lowest. The personality test papers of these individuals were removed from the rest of the pile, and the responses made by high and low groups to each item were tabulated. On the Minnesota Test, where the respondent marks each item 1, 2, 3, 4, or 5, the mean value of the responses in each group was obtained, and the difference between means of high and low groups checked for significance by Fisher's "*t*." On the Thurstone scales, where the response consists of checking or not checking each item, simple percentages were obtained and checked for statistical significance. By this method, personality items which were clearly related to interest scores were sorted out from those which were not.

The next step, for tests in which there turned out to be enough of such items to make the task worthwhile, was to rescore the papers by a key which included only the discriminating items. The correlation of interest score with this special score based on selected items furnishes a better estimate of the relationship of interest and personality factors than does the original correlation. For the most significant relationship obtained in this way, a further check was made. A group of 38 cases, not used in the item analysis, was chosen from the author's counseling files. The interest-personality correlations were calculated for this group also.

One problem arose in the design of this investigation, the solution of which was not entirely satisfactory. If the women's form of the Strong Blank were used for women subjects, group scores could not be obtained. If the men's blank were used, there would be some question as to the meaning of the women's scores. Since there is considerable evidence that meaningful scores for women can be obtained with the men's blank, we decided to use it, watching for any indication of sex differences in the results. Unfortunately, only about a third of the subjects were men, so that the most unequivocal results rest on the smaller number of cases.

Results

Male Subjects. Correlations between interest scores and other personality variables are shown in Table 1, which reveals several interesting facts. First, the *Social Adjustment* test is the only one from the *Minnesota Personality Scale* which shows more than one significant correlation, with Strong scores. The *Family* and *Emotional* tests do not correlate with anything. *Morale* is negatively related to *Verbal* interest scores and *Economic Conservatism* is negatively related to *Human Science* interest scores. Otherwise, correlations involving these two sub-tests are too low for significance. Second, correlations of the *Social* score with Groups I and II, on the one hand, and Group IX, on the other, are opposite in sign and approximately equal in size. Third, one or both of the religious attitude scales correlate significantly with all Strong scores except Group X. The directions of these relationships are consistent. Less religious attitudes go with science scores, more religious with personnel and business.

Table 1

Correlations between Strong Group Scores and Personality and Attitude Scores, $N = 55$ Men

		Personality Tests						
		Morale	Social	Family	Emot.	Ec. Cons.	Attitude God	Attitude Church[1]
Interest	Group I Human Science	−.24	−.32*	.00	−.14	−.20*	−.26*	.34†
	Group II Technical	−.03	−.35†	.27	−.06	−.01	−.32*	.32*
	Group V Personnel	.17	.25	−.12	−.12	−.00	.04	−.20*
	Group VIII Office	.23	.14	−.03	.00	.24	.17	−.20*
	Group IX Sales	.03	.40†	.09	.22	.00	.38†	−.25
	Group X Verbal	−.34†	.04	−.23	−.10	−.10	−.02	.23

[1] High scores indicate *unfavorable* attitude.

* Significant at 5% level (Fisher's "*t*" test).

† Significant at 1% level.

Item analysis of the relationship between Group X and Morale turned up very little that was interesting. Only three items gave significant "*t*'s." These are: 2. Joys of family life are much over-rated; 9. A high school education is worth all the time and trouble it requires; and 39. A good education is a great comfort to a man out of work. The correlation in this case seems to reflect simply a slight tendency for men with high verbal interest scores to be less optimistic about things in general. Practically all the items showed this trend, but in none except these three was it marked enough to result in a significant "*t*" (5% level). A pessimistic slant in Group X men may be attributed to either the higher intelligence or the larger number of dislikes on the Strong blank itself which is known to characterize many men in this group.

Similarly, item analysis of the relationship between Group I scores and *Economic Conservatism* was unfruitful. Only two items showed significant differences: 186. If our economic system were just, there would be much less crime; and 193. A man should be allowed to keep as large an income as he can get. Here too most of the other items showed a tendency for individuals with high Group I scores to respond less conservatively, but differences were slight. Such a trend may again reflect nothing more than the slightly superior general intelligence usually associated with high Group I scores.

It was in the item analysis of the *Social* subtest that the most interesting findings appeared. Thirteen items out of the sixty-five showed significant differences between high and low groups on one or more of the three interest scales with which the scores were correlated. A scrutiny of these items indicates that they have in common one rather limited, specialized phase of an individual's social personality—namely, his attitude about the desirability of friendship with *many people.* The four items on which differences are most pronounced for all three interest groups under consideration are: 55. Do you like to meet new people?; 72. Do you get along as well as the average person in social activities?; 84. Do you like to know a great many people intimately?; and 88. Do you prefer to participate in activities leading to friendships with many people? To each of these questions, the sales group gives a much more social answer than the two science groups. Items concerned with other attitudes toward people and social affairs do not differentiate. Self-consciousness, embarrassment, unhappiness about one's blunders, lack of aggressiveness, failure to get along with people—none of these has any demonstrable relationship to the interest factors differentiating salesmen from scientists. The social characteristic involved here is a matter of *preferences* rather than *adjustment.*

When the *Social* test was scored by a key made up of only the 13 differentiating items, the results shown in Table 2 appeared. There seems to be little question that the social trait tapped by the 13 discriminating items is related more closely to scientific and sales interests than is social adjustment in general. The most conclusive coefficients are those based on the cases from the files. While the Group I r is lower

Table 2

Correlations between Strong Scores, Groups I, II, and IX, and Scores on Scales for Measuring Social Adjustment or Attitudes

	Complete Social Scale (Orig. Group) $N = 55$	13-Item Scale (Orig. Group) $N = 55$	13-Item Scale (Cases from Files) $N = 38$
Group I	−.32	−.57	−.44
Group II	−.35	−.56	−.52
Group IX	.40	.58	.60

for this group than for the original subjects, Group II is about the same, and Group IX is actually higher. The full importance of these correlations becomes apparent when we realize that on the basis of the 13 social items alone we could predict Strong scores on Groups I, II, and IX with about as much accuracy as we predict school marks from intelligence test results.

The item analysis of the *Attitude toward God* scale showed, as one might expect, that the principal differences stemmed from a tendency of the scientifically-minded to avoid the more mystical positions, such as "God is the underlying reality of life." There was no difference on anti-religious items; few if any of these subjects checked them. On the *Attitude toward the Church* scale, the principal difference between scientists and personnel and business men seemed to be a greater tendency toward skepticism among the former about whether or not the church produces all the good effects claimed for it. Differential responses to such items as "I feel that church attendance is a good index of the nation's morality" are typical of this trend. On this scale also, the anti-religious items were avoided by everybody. The difference between scientists and business-personnel men might be described as a difference in amount of indifference, not opposition, to the church.

The last step of the procedure, scoring the test on discriminating items alone and recalculating the correlations, was carried out for the *Attitude toward the Church* scale. (There were too few items in the other to make it feasible.) The outcome is shown in Table 3. Again the correlations are larger with the scores based on discriminating items alone, but the increase is less marked than for the social variable. Since on scales of this type subjects check only those items with which they agree, it is probable that irrelevant items carry less weight in the original score than they do where the other scoring technique is used.

A possibility that a few items on the Strong test that happened to be closely similar to the social items we have sorted out might account for the obtained correlations also needed to be checked. An item analysis in the reverse direction was used for this purpose. That is, the responses on the *Strong* test of the 20 individuals who were highest on the 13-item social scale were tabulated and compared with the responses of the 20 lowest. Differences showed up on a large number of interest items scattered throughout all parts of the blank in such diverse areas as, for instance, liking geography and handling horses. This suggests that a different general outlook differentiates these people and not just a few specific preferences.

Table 3

Correlations between Strong Scores, Groups I, II, V, VIII, and IX, and Scores on Scales for Measuring Attitudes toward the Church. $N = 55$

	Complete Att. Church Scale	10-Item Scale
Group I	.34	.46
Group II	.32	.40
Group V	−.29	−.28
Group VIII	−.29	−.40
Group IX	−.25	−.30

Another question of some significance is whether or not social and religious attitudes are correlated. This seems to be the case. The correlations between the special social scale and the special church scale is .40. Thus interest scores could be predicted from a combination of the two with only slightly more accuracy than from the social score alone.

Female Subjects. For the 122 women in the study, correlations between Strong scores and other personality variables are shown in Table 4. There are more statistically significant r's than for the men's group because the larger number of cases makes a smaller coefficient significant. Numerically, most of the coefficients are somewhat lower than those in the corresponding cells of Table 1. It is interesting to note that all the significant relationships are in the same direction for women as for men. The principal difference between Tables 1 and 4 is with regard to the religious attitude scales. In Table 4 only one of the relationships between religious attitudes and interests is significant and that one (Group V vs. *Attitude toward God*) is very low. Group V interests, which for men are most closely associated with favorable attitudes toward the

Table 4

Correlations between Strong Group Scores and Personality and Attitude Scores. $N = 122$ Women

		Personality Tests						
		Morale	Social	Family	Emot.	Ec. Cons.	Attitude God	Attitude Church
Interest	Group I Human Science	−.25**	−.27**	−.15	−.20*	−.12	.01	.06
	Group II Technical	−.14	−.27**	−.08	−.07	−.11	−.17	.14
	Group V Personnel	.23**	.34**	−.01	.14	−.19*	.19*	−.05
	Group VIII Office	.19*	.05	.14	.07	.08	.04	−.06
	Group IX Sales	.16	.38**	.13	.12	.23**	.08	−.15
	Group X Verbal	−.09	−.01	−.07	−.07	−.06	−.05	.11

* Significant at 5% level (Fisher's "t" test).

** Significant at 1% level.

church, are in women most clearly related to social characteristics and morale.

Item analysis of scales other than the *Social* again selected only scattered items, difficult to fit into a coherent pattern. But out of the 53 items on the Social test (Women's Form) 41 differentiated high from low people on one or more of the interest scales. A great many of these are significant for Group V only. Women high in personnel interests give more extraverted responses than others to almost all types of question. A smaller set of 19 items differentiated scientific and sales interests, as did the set in the men's data. There is more of a tendency, however, for items pointing toward social maladjustment to differentiate between these interest groups of girls. For instance, the item, "Do you find it easy to make friendly contacts with members of the opposite sex?," has no differentiating value for men, but shows up significantly in the analyses of four scales for women (Groups I, II, V, and IX), the more social answer characterizing girls with personnel and sales interests, the less social, girls with science interests. The results of rescoring the blanks by the 19-item key and recalculating correlations are shown in Table 5. Coefficients are increased by eliminating non-discriminating items, but not so markedly as is true for men. This also may indicate that items having to do with social *adjustment* and happiness are less irrelevant to interests for women than for men and thus have a less depressing effect upon correlations. Unfortunately, relationships in the women's data could not be checked in a new group, as the counseling files contained too few cases of women who had taken the men's Strong test.

Table 5

Correlations between Strong Scores, Groups I, II, and IX, and Social Scores on Minnesota Personality Scale. $N = 122$ Women

	Complete Social Scale	Special 19-Item Scale
Group I	−.27	−.41
Group II	−.27	−.37
Group IX	.38	.47

Discussion

The facts outlined above can be brought together under a few main headings. First, it is to be noted that interest scores on Group I, Group II, and Group IX scales are the only ones showing consistent significant relationships to other personality factors. We have uncovered nothing of any importance about the other group scales, except that women's Group V interests tend to accompany higher socialization and morale and men's Group V and Group VIII interests go with a relatively favorable attitude toward the church. The scientific and sales interests are at opposite poles in all the relationships investigated in this study. Many of Strong's reported results show the same trend (4).

The second generalization is that there is a social factor related to interest differences. This factor, for men, is centered around the sort of social stimulation they prefer and seek. The more a man tends to avoid large numbers of acquaintances and indiscriminate social affairs, the more likely he is to show the interests of scientific men. The more satisfaction he takes in social affairs involving large numbers of people, the more likely he is to resemble salesmen in his interests. Items having to do with maladjustments and unhappiness do not differentiate between men's interest types. Women show the same general sort of difference between persons with scientific and sales interests, but the nature of the social factor is a little less clear-cut. The same items concerned with liking or avoiding social affairs constitute its important component, but some other less-easily-catalogued items also differentiate. Girls with high scientific interests are likely to be somewhat less well-adjusted to the opposite sex, somewhat less confident and happy in their social relationships than girls with other types of interest. This may be because in the direction their interests are taking they are departing rather markedly from what society expects of women and are thus introducing some strain into their relationships with people. It is impossible to determine from the data at hand which of these variables is cause and which effect. A genetic study should throw light on these doubtful issues.

Postulation of an important social factor around which other attitudes cluster suggests hypotheses explaining several puzzling facts turned up from time to time in interest research. Many have wondered, for instance, why artists should be classified with physicians, psychologists, architects, and dentists, by the correlations between interest scores. A similarity in social outlook might be the basis for these correlations. The predominance of scientific interests in adolescent boys is another such research finding. The social attitudes of 15-year-old boys, if they carry other attitudes along with them, could account for this. The correlation of interest scores with success in selling life insurance might also involve this third factor. If the high-scoring men have a different social outlook, that could easily account for their selling more insurance. All these hypotheses could be checked statistically without a great deal of trouble.

Another general finding from this study seems worthy of mention,—the general *lack* of relationship between interest scores and what we loosely term "neurotic tendency." There is no correlation for either sex between interests and family adjustments. There is only one low significant correlation between interests and emotional adjustment, and this shows no particular trend in the item analysis. None of the items on the social scale referring to relatively serious maladjustments or neurotic symptoms differentiate in any interest group. It may well be that there are two independent variables in personality, the direction which it takes and the success or effectiveness of the adjustment made. We have perhaps given too much attention to adjustment in personality study, and too little to this other consideration of the direction of development. There is certainly no evidence here to suggest that neurotics tend to pile up in certain occupations.

That correlations with religious scales are significant for men but not for women may indicate that vocational interests are a more fundamental factor in male than in female personalities in our culture, more closely integrated with the individual's whole cosmic orientation. As in most studies, our women subjects averaged slightly more religious than the men, but the differences were not statistically significant. It is to be remembered that the differences in men are not on the anti-religious items but on those showing skepticism. The tendency one might expect to find for business men to be more conservative than scientists shows up more plainly with regard to religious issues than in regard to economic opinions.

Finally it should be noted that the results presented in this paper coincide in several respects with some other exploratory studies. The table given by Darley (2) shows the largest number of significant critical ratios for the scientific interest groups and the social scales, particularly *Social Preferences*. Sarbin and Berdie (3) find more significant differences in Allport-Vernon scores for Groups I and II than for the others. Tussing (5) finds that scoring keys for the Strong blank can be constructed to measure somewhat the same factors as the Bernreuter F-1C and F-2S scales and the Bell Social scale, but not the other Bell scales. It would be possible to fit the results of this study into Bordin's (1) theory that "in answering a Strong Vocational Interest Test, an individual is expressing his acceptance of a particular view or concept of himself in terms of occupational stereotypes," if we assume that attitudes toward social affairs and friendships are important features of the stereotypes for some occupations.

Summary

1. A social factor having to do with feelings about organized social affairs and friendships with many people is related significantly to Group I, Group II, and Group IX scores on the Strong Vocational Interest Blank.

2. Religious attitudes show a moderate correlation with vocational interests in men but not in women.

3. Scores indicating neurotic tendencies show no appreciable relationship to any kind of interest scores.

Received December 1, 1943.

References

1. Bordin, E. S. A theory of vocational interests as dynamic phenomena. *Educ. Psychol. Meas.*, 1943, 3, 49–66.
2. Darley, J. G. A preliminary study of relations between attitude, adjustment and vocational interest tests. *J. Educ. Psychol.*, 1938, 29, 467–473.
3. Sarbin, T. R., and Berdie, R. The relation of measured interests to the Allport-Vernon study of values. *J. appl. Psychol.*, 1940, 24, 287–296.
4. Strong, E. K., Jr. *Vocational interests of men and women.* Stanford University Press, 1943.
5. Tussing, L. An investigation of the possibilities of measuring personality traits with the Strong Vocational Interest Blank. *Educ. Psychol. Meas.*, 1942, 2, 59–74.

A Worry Inventory

A. H. Martin

The University of Sydney, Sydney, Australia

This inquiry originally constituted an endeavour to supply an alternative form to the usual questionnaire used as a measure of maladjusted trends in personality. In place of presenting questions requiring the underlining of answers either "yes" or "no," direct but brief description of symptoms were set out. Test subjects were first required to underline those items which had ever worried them. When this task was completed, they were further required to make a ring around the number of any of the underlined items which at present still constituted a cause of worry. Two scores were thus obtained; the first consisted of an "effective" score of worry items; the second score noted existent present worries. In administration this inventory proved to be far easier for candidates to answer than the usual questionnaire form of personality inventory.

The questionnaire generally used for vocational guidance work in Sydney, Australia, for candidates from fifteen to twenty-five years of age is Thurstone's[1] selected questions, modified by a few necessary local emendations and the addition of three "jokers," i.e., catch questions to which the better answers are "Yes."

Regularly each year some undergraduates have commented on the difficulty of following the directions for this questionnaire. The effort of fine discrimination required to arrive at honest decisions, appears to be great. At the same time, the questionnaire does actually sort out many cases showing inferiority trends, shyness or "nervousness." These individuals may be directly helped by an interview or two involving a brief analysis of the subjects' self-centered attitude. There were 42 questions in the original Sydney inventory and the average number of "yields" per person was from 10 to 11 questions. Self-confident individuals of a good "sales" type tend to register from 3 to 6 yields. Occasionally types with pronounced "inferiority" or neurotic trends show from 15 to 35 question yields. A Factorial Analysis[2] shows the questions themselves tend to cluster about such trends as schizoid, submissive, a-social and manic-depressive.

---

[1] Thurstone, L. L., and Thurstone, T. G. A neurotic inventory. *J. soc. Psychol.*, 1930, 1, 3–30.

[2] Gibb, C. A. Personality traits by factorial analysis. *Aust. J. Psychol. Phil.*, June, 1942, 1–15.

Worry Inventory

Since the general principles involved in the Thurstone type of questionnaire had proved so useful for many years, it seemed advisable to retain the principle involved and yet to seek some simplification of the questionnaire method. Accordingly the direct method was decided on for a try-out. Every existing list of worries, patterned after Woodworth's[3] original lead, tends to be cast in the form of questions. But this method is neither economical in its setting out nor is it often easy for the candidate to answer. The itemized list, used here, however, directly denominates each symptom, which has been reduced as far as possible to very simple and unambiguous terms. The method of marking the items was first to get subjects to underline those which had ever caused them worry. After completing the list thus, they were required to put a ring around the number of each item which was still a present cause of worry or distress. Two individual scores were thus secured. If the items are honestly marked, the first indicates the general nature of individual worries; the second score is indicative of the nature and extent of existent worries. This procedure of underlining and ringing was borrowed from the method prescribed in the Pressey Emotional X.O. Test.

The Group Tested

The original group of individuals tested consisted of undergraduates attending classes in Psychology in years I, II or III in the University of Sydney, together with a few outside persons attending one of these single courses. About fifty per cent of the undergraduates were students enrolled for evening courses. Altogether 100 persons were used as subjects, 48 being females and 52 being males. Their ages varied from 17 to 47 years, but gave a preponderance to subjects of earlier years with a very decided skew towards the higher years. The average age for both sexes combined was 20.4 years. In order to enable observations to be made upon age trends, the scores were separated into male and female results under the following age divisions: (a) 17 to 19 years, (b) 20 to 21 years, and (3) over 21 years. No differentiating trends in age or sex appeared, either in yields to particular types of items, in actual numbers or in the proportion of existent worries to general worry problems. The six subgroups, though fairly uniform in trends, were not large enough for any reliable inferences to be drawn. Some degree of selection must have been operative in the male section with respect to military service, hence this sampling can hardly be regarded as "average" and fully representative. In the female section no such factors of selection probably came into play.

---

[3] Franz, S. I. *Handbook of mental test methods.* New York: Macmillan Co., 1920.

The subjects did not record their names on the sheets unless they wished, but merely indicated sex and age.

The Present Inventory

The items with their two results given alongside, are shown in Table 1. The items were arbitrarily selected by the writer on a representative

Table 1

Items in the Worry Inventory with Percentage of "Yields" for 100 University Students

	Percentage of Yields	
	I*	II**
1. Poor health	20	7
2. Being different in appearance to other folk	23	7
3. Having a poor appetite for food	3	0
4. Bad indigestion	3	1
5. Lying awake at night	26	9
6. Disagreeable dreams or nightmares	23	6
7. Constant aches or pains	2	0
8. Night sweats	3	0
9. A dislike for certain kinds of food	21	11
10. Spells of dizziness	14	6
11. Thoughts of death	28	6
12. Being nervous or shy	75	18
13. Sudden heart tremors or palpitations	11	7
14. A tired feeling after waking	29	19
15. Frequent headaches	20	8
16. Entertaining folk	34	16
17. Thoughts of suicide	9	0
18. Being found fault with	48	16
19. Not taking part in sports or games	15	1
20. Sleep-walking	3	0
21. Being too quarrelsome	13	2
22. Loneliness	26	8
23. Bad asthma or shortness of breath	5	3
24. Feeling bored or "fed up" with life	29	4
25. Sexual problems	31	12
26. Mind wandering or day dreams	25	6
27. Constant bad luck	3	0
28. Being teased or made a fool of	30	7
29. Not being understood or appreciated by people	16	2
30. Meeting members of the opposite sex	30	6

Table 1—*Continued*

	Percentage of Yields	
	I*	II**
31. Being treated unfairly by others	30	5
32. Leaving tasks unfinished	20	12
33. Getting no pleasure out of life	8	2
34. Talking too much	15	7
35. Lack of success in your studies	22	6
36. Not being loved at home	6	1
37. Addressing groups or meetings	38	21
38. Being closely watched or observed	30	12
39. People's wickedness	7	5
40. Being a failure in your job	10	3
41. Feeling self-conscious	54	27
42. Brooding over your sins	9	1
43. Having wrong thoughts	23	8
44. Going into tunnels or subways	2	1
45. Not being your real self	10	3
46. Some speech difficulty such as stammering or lisping	10	6
47. Looking down from a height	31	16
48. The need to do things over and over	10	2
49. Things seeming unreal to you	11	3
50. Talking in your sleep	7	3
51. Not being able to converse easily with people	33	13
52. Making one of a crowd	10	2
53. Your lack of true friends	11	5
54. Blushing	37	16
55. Lack of self-confidence	34	17
56. Twitching of the face, head, body or limbs	7	4
57. Your religious beliefs	10	0
58. Some vague constant fear or worry	14	6
59. A fear of going insane	7	1
60. Inability to fix your attention on books, work or studies	28	18
61. Having a bad temper	21	5
62. Being afraid	14	4
63. Meeting with or talking to elders or social superiors	20	7

---

* *Directions I.* First read carefully through every item mentioned below and underline *everything* which has ever troubled or worried you.

** *Directions II.* Now go through each item underlined and if it is still worrying you, draw a ring around its number.

basis from a wide array of existing "personality" questionnaires. Where two items were not mutually exclusive and tended to cover a particular symptom, the better, in his judgment, was retained.

Of all the sixty-three items only seven proved to be unproductive by showing a total yield of only 3% or less. These items are Nos. 3, 4, 7, 8, 20, 27, and 44. The last indicated that claustraphobia is a fairly infrequent symptom compared with No. 47, which involves "looking down from a height." For use with a normal and representative group of individuals the seven items could well be omitted.

The remaining fifty-six questions were arbitrarily distributed by the writer under the headings shown in Table 2. They are arranged in descending order of frequency of occurrence.

Table 2

Grouping of Items in the Worry Inventory by Categories

Type of Difficulty	No. of Items	Average Percentage		Item Numbers
		Ever Worried	Still Worried	
Sex	3	31	9	25, 30, 48
Inferiority	21	28	5	2, 12, 16, 18, 22, 28, 29, 31, 34, 35, 36, 37, 38, 40, 41, 46, 51, 53, 54, 55, 63
Physical Symptoms: Conversion	10	16	8	1, 9, 10, 13, 14, 15, 23, 44, 47, 56
Schizoid Trends	16	18	5	5, 6, 11, 17, 24, 26, 32, 33, 45, 48, 49, 50, 52, 58, 59, 60
Religious Difficulties	3	12	5	39, 42, 57
Fear and Anger or Cyclothymic Trends	3	14	3	21, 61, 62
Total	56	21	5	

Psycho-analytic literature would probably technically label these as erotic, narcissistic, conversion or hypochondriac, repression, super-ego and accumulations of id symptoms and trends. On the other hand, the nomenclature of various factor analysts[4] would use other specific terms such as "lack of confidence," "solitariness" for the inferiority and schizoid trends. The remainder would hardly find a place. The chief concern of the investigation was to indicate certain practical advantages of the method of presentation and the method of marking the items.

The results presented exhibit a decidedly different aspect to those obtained from the general run of such inventories, both in the matter of items, the item groups and the extent of the prevalence of such difficulties. This inventory, therefore, tends to shed a different light on the nature of nervous worries and problems to that of the ordinary personality questionnaire. It cannot, of course, be indicated to what extent the disclosures of such worries was due to the anonymity covering the individuals who answered the inventory, but at least one-third of the group voluntarily subscribed their names. Further the writer was brought into contact with the groups concerned as lecturer, so that a certain degree of rapport possibly existed.

---

[4] Gibb, C. A., *op. cit.*

Do these average frequencies really indicate the relative importance of the groups of items? One cannot answer "yes" directly to this question. In the first place the number of questions relating to "Inferiority" amounted to 21 or almost 40 per cent of the total of the items, yet the group ranks second in average importance to that of sex difficulties which is represented by only three items.

But seven items of the inferiority group approach or show a greater percentage of yields than the most heavily weighted item, No. 30 of the Sex Difficulties group. There is, hence, no conclusive evidence here supporting either the exclusively Freudian or the purely Adlerian theories of personality. The individual reader must interpret the results in his own way. It can be stated, however, that the present inventory is possibly more "revealing" in many ways than the Thurstone Questionnaire.

Later on, both the Worry Inventory and the Thurstone Inventory were administered to a group of 36 students of both sexes, for the sake of comparison. The following correlations were obtained:

Thurstone Inventory and Worry Inventory Underlining.............. = .81 ± .038
Thurstone and Worry Inventory Ringing.............................. = .75 ± .056
Worry Inventory Underlining and Worry Inventory Ringing.......... = .81 ± .038

To supplement this a test of extraversion was administered to the same group within a week of giving the others. The results were:

Extraversion and Thurstone.................................... = − .40 ± .035
Extraversion and Worry Inventory Underlining................... = − .56 ± .077
Extraversion and Worry Inventory Ringing........................ = − .67 ± .06

Thus, there appears to be a remarkably consistent constant factor emerging from all these tests as revealed by these correlation results which is probably that of intro-extraversion. While its symptoms are legion, one underlying cause, that of ego-centricity, appears as the prime cause of the human difficulties.

Summary

1. In the present "Worry Inventory" a list of sixty-three items was presented to a group of 100 university students of both sexes ranging

in age from 17 to 45 years. The inventory is intended to be representative rather than exhaustive.

2. The method of administration of the Inventory proved simple and far less difficult to answer than the usual list of personality questions.

3. Its use tended, possibly under a cloak of anonymity, towards a higher yield of nervous difficulties and a wider field of symptoms, than is usually uncovered by the usual questionnaire method. However, some fair proportion of the excellent results is probably due to the direct approach used in the present Inventory.

4. The Worry Inventory correlated to a very high degree with the Thurstone Questionnaire.

Received November 20, 1943.

Social Factors Annoying to Children

Rose Zeligs

Avondale Public School, Cincinnati, Ohio

The present world upheaval adds to the bombardment of disturbances and annoyances that affect children and contribute to instability and lack of poise. A knowledge of annoying situations in children's daily experiences should enable parents and teachers to eliminate many of the disturbing factors and to help the children adjust to annoyances that are unavoidable.

This paper will discuss children's reactions to annoyances they often experience. The subjects comprised 145 sixth-grade boys and 140 girls from two suburban elementary public schools of Cincinnati. Ninety-nine per cent of the children were native born, 70 per cent Jewish, 27 per cent were non-Jewish, and 3 per cent were colored. Their average chronological age was 11 years and 8 months, and their average mental age, according to the Otis Group Intelligence Scale, Advanced Examination, Form B, was 14 years. The socio-economic background was a little below very high on the Sims Score Card.

These 285 sixth-grade children were asked to list all things which annoy, irritate, and bother them. The items were classified under social relationships, health and appearance, home and family, school, hobbies and interests, foods, personal conduct, games and amusements, fears, inconveniences and annoyances, animals and insects, and environmental conditions.

The following year the material was presented to the 285 sixth-grade children of the same schools, described above. The material, in the form of separate tests for boys and girls, provided for five different degrees of feeling toward every item listed. These were *like, don't mind, don't like, hate,* and *hate much*. The children were asked to encircle the expression that best described their feeling toward each item listed. The material was tabulated, changed to per cents, and arranged according to frequency for *hate much*.

This paper is a report of the items classified under social relationships, home and family, school, and personal conduct.

Attitudes of Boys

Table 1 lists social situations that are annoying (don't like, hate, and hate much) to 65 per cent of the boys studied. They are especially

Table 1

Attitudes of 145 Boys Toward Social Situations

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
Be blamed for something I didn't do	59	To be slapped	34
People who cheat	59	To be ridiculed	33
Unfair things	59	Sarcastic remarks	33
A bully	56	Two against one	28
War	56	People who talk too much	28
Have people angry at me	51	Ugly people	26
To lose someone's things	46	See girls act grown up	26
To be poor	45	See people crying	26
People who show off	44	Hear talk on self-control	24
A sissy	44	To borrow money	22
People who tell lies	43	Old maids	22
People who laugh when I get hurt	43	People who talk too fast	22
People who always argue	40	People who talk too slowly	21
Stupid people	40	To fight	19
To lose in a fight	37	Long religious sermons	18
To fight little boys	37	To be alone	18
A mean person	37	People on diets	14
To get into trouble	37	To talk on telephone	13
Tattletales	35	Certain boys	13
		Certain girls	13
		People who act funny	12

The average distribution for all items, in terms of per cent, was: Like 16, Don't mind 19, Don't like 21, Hate 17, and Hate much 27.

disturbed when blamed for something they did not do and by people who cheat, unfair things, a bully, and war. They do not like to have people angry at them, to lose someone's things, or to be poor. People who talk too much, who lie or argue, who are stupid, mean or sarcastic, who laugh when others get hurt are very much disliked by the boys. Sissies, show-offs, or tattletales rate low with these boys. They do not like to lose in a fight or to fight with little boys, to get into trouble, to be ridiculed, or to be slapped. The boys vary in the number of situations that annoy them and in the degree to which they are annoyed.[1]

Table 2 gives home and family situations that are annoying to many boys. They dislike most to get a whipping, especially for something they did not do, to be scolded, punished, or nagged.

Some of the boys do not like to do various chores around the house, run errands, or take care of smaller siblings.

---

[1] Rose Zeligs, The relationship of emotional and personality traits to learning in children (unpublished Doctor's dissertation, University of Cincinnati, 1937).

Table 2

Attitudes of 145 Boys Toward Home and Family

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
To get a whipping for something I didn't do	62	To wash and dry dishes	13
To get a whipping	50	To sweep floors	13
To be scolded	42	To wash the bathtub	13
Let brother break my games	36	To work around the house	13
To be punished	36	To go to bed late	12
To be nagged	35	To dust furniture	12
To clean the cellar	23	To set the table	12
To sew	21	To fight with brother	10
To wait in store while mother shops	20	To clean gold-fish bowl	10
My sister's singing	18	To cook	10
To go to bed early	16	To go to the store	10
To scrub	16	To take care of house	9
To clean the house	14	To shovel snow	9
To take care of cousin	13	To go to bed	9
		To rake the yard	8
		To run errands	6

The average distribution for all items, in terms of per cent, was: Like 23, Don't mind 33, Don't like 16, Hate 11, and Hate much 17.

Table 3

Attitudes of 145 Boys Toward School Situations

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
To get low marks	40	To go to Sunday School	13
Certain teachers	40	To do nightwork	13
Teachers who have pets	36	To have to miss school	12
Be sent to principal	36	To write book reports	11
To hear teachers preach	36	To study grammar	11
Teachers who dislike me	36	To go to school	9
Writing lessons	23	To give talks	9
To memorize poems	21	To study history	8
High singing	18	To read my compositions	7
To write compositions	17	To take art	7
Singing in school	14	To take tests	6
Arithmetic	13	To study geography	4

The average distribution for all items, in terms of per cent, was: Like 34, Don't mind 27, Don't like 14, Hate 9, and Hate much 16.

Though many school contacts are enjoyed by the boys, it can be seen from Table 3 that the boys dislike most to get low marks, certain teachers, teachers who have pets, who preach, or who do not like them, and to be sent to the principal. Almost 20 per cent of the boys hate or hate much to go to school while 40 per cent of them have that same reaction towards Sunday School. Writing and singing lessons represent the least-liked subjects, while manual training and gymnasium work are the favorite subjects.

Table 4 lists the boys' attitudes toward personal conduct. Heading the list are character and personality qualities. To curse, tell lies, or be accused of lying and to have a guilty conscience or bad habits are greatly disliked. Other annoying situations involve the loss of money, fountain pen, or other possessions, or to lose in games.

Table 4

Attitudes of 145 Boys Toward Personal Conduct

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
To curse	57	To break windows	38
To tell lies	52	To brag	36
To be stupid	51	To get angry	32
To be accused of lying	51	To lose in a ball game	30
A guilty conscience	47	To lose in any game	28
To lose money	46	To hurt someone	28
To cry	45	To lose in card games	26
To lose a fountain pen	45	To lose in marbles	25
To bite my nails	41	To attract attention	18
To lose anything	41	To kiss people	15
To have bad habits	38	To be still	12

The average distribution for all items, in terms of per cent, was: Like 4, Don't mind 13, Don't like 25, Hate 22, and Hate much 36.

Attitudes of Girls

Girls express annoyance to a greater number of situations and to a greater degree than do the boys. Table 5 contains many more items showing girls' dislikes in social situations than are found in Table 1 for boys. Most disturbing to the girls is to witness pain and suffering, people getting killed, sickness and death. Social relationships that result in lowering their dignity or personal status are very annoying to the girls. These include being accused of something they did not do, or of lying, being called a cheater, made fun of, treated like a baby, called names, scolded, or called a poor sport.

Table 5

Attitude of 140 Girls Toward Social Situations

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
See people getting killed	68	A stale joke	36
Be accused of something I didn't do	68	Grouchy children	36
Be called a cheater	61	Selfish people	36
See people suffer	56	Be bothered when reading	35
Be treated like a baby	56	Quarrel with boy friend	35
Be made fun of	55	Children acting babyish	33
Hear someone has died	53	When friends act foolish	33
Be called names	51	People who talk too much	33
Be scolded	50	People who don't like to do anything	33
Be bothered or pestered	50	A dull party	32
A bully	48	People who gossip	32
People who are fakers	48	To lose in a game	32
To see someone sick	48	Thoughtless people	32
To be called a poor sport	48	*Play with someone I don't like*	31
To be accused of lying	48	Not to go to parties	31
Associate with dirty people	47	People who talk in movies	31
People who lie	47	To be teased	31
Nasty people	47	Visits to boring people	31
People minding my business	46	Be teased about boy friend	28
Jealous people	46	Discourteous children	27
A dry speech	46	Silly people	27
To be embarrassed	45	To *fight*	26
To see a child whipped	45	To get angry at someone	25
Sissies	44	To be patronized	24
Children who cheat	43	Smart person acting dumb	23
An untrue friend	43	People who are cross	23
To be insulted	43	Not to go to dancing school	23
To play with mean children	43	People who are rough	23
Constant nagging	43	Persisting people	22
See people biting their nails	43	To be disappointed	22
Conceited people	42	Kissing games at parties	21
Stubborn people	40	To pay full car fare	22
Have privileges taken away	40	To visit sick adults	20
People who are pests	40	Talk about self control	20
People who brag	40	Boys who make love to me	19
When friends get sick	40	Talk long on one subject	19
See a crippled person	39	To take revenge	18
See children show off	39	People who act funny	17
Sloppy people	38	Have big responsibilities	16
A hairy actor	38	To talk long on phone	15
Stupid people	37		
To be yelled at	36		

The average distribution for all items, in terms of per cent, was: Like 14, Don't mind 14, Don't like 20, Hate 22, Hate much 30.

Most of the girls express extreme dislike for antisocial people such as those who lie, mind other people's business, are jealous and nasty. They don't like sissies, children who cheat, and an untrue friend. Personality traits annoying to the girls are conceit, stubbornness, sloppiness, grouchiness, foolishness, selfishness, stupidity, and thoughtlessness. Constant nagging and teasing, visits from boring people, and individuals who bite their finger nails, talk in movies, and gossip are disliked. In general, the girls reflect very definite standards of social conduct and an unfavorable attitude toward those who do not live up to those standards.

Also, in their attitudes toward home and family the girls reflect standards of family relationships. As seen in Table 6, practically all the girls dislike to make their parents unhappy or angry, to be punished or scolded. Though they do not like to be pestered by siblings, most of the girls like their little sisters and brothers and do not mind taking care of them. Practically all the girls like babies. Some of them dislike housework, but very few girls dislike company or going out on Sunday.

Table 6

Attitudes of 140 Girls Toward Home and Family

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
Make parents unhappy	54	Not to see relatives	21
A spanking	52	To wash and iron clothes	20
To irritate my mother	50	To be bothered by sister	20
When mother is angry	41	To scrub	16
To be punished	41	To wash stockings	12
To be scolded	37	To wash dishes	12
To stay home on Sunday	36	To dust furniture	11
A dirty yard	36	To make beds	11
To see little brother cry	31	My sister's singing	10
To have my things bothered	28	To sweep	9
Not to help mother	24	To darn socks	9
When mother won't buy things I want	23	To go to bed late	9
To be bothered by brother	22	To clean the house	7

The average distribution for all items, in terms of per cent, was: Like 26, Don't mind 24, Don't like 18, Hate 14, and Hate much 18.

As shown in Table 7, maladjustment in school is displeasing to the girls. They are extremely disturbed by bad report marks, unsatisfactory lessons, or poor conduct in school, especially if these require their mothers to come to school. They do not like teachers who get angry, scold, or give long lectures, or fail to mark papers. Writing lessons are liked least

Table 7

Attitudes of 140 Girls Toward School Situations

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
Bad report marks	58	Writing lessons	21
When mother has to come to school	40	To have nightwork to do	21
Not to know my lesson	39	Teachers who don't mark papers	19
To be bad in school	36	To discuss poems	18
Not to have my homework	34	To copy things over	18
To make teacher angry	33	To memorize poems	18
When I can't answer a question	33	To make book reports	17
To get nervous in a test	31	To go to Sunday School	13
Too much nightwork	30	To carry books	13
To do hard lessons	27	To study	13
Long lectures by teachers	25	To have tests	11
To talk out in school	24	Certain teachers	11
Teachers who scold	23	To have school over	10
To memorize lessons	22	To write compositions	9
To stay home from school	22	Grammar	6

The average distribution for all items, in terms of per cent, was: Like 20, Don't mind 23, Don't like 17, Hate 16, and Hate much 10.

of all the subjects, while gymnasium work, spelling, drawing, and geography are liked most.

Antisocial character and personality traits (Table 8), as expressed in their personal conduct, are annoying to the girls. Telling lies, being bad, mean, or unreliable, doing wrong, or getting into trouble are marked

Table 8

Attitudes of 140 Girls Toward Personal Conduct

Item	Per Cent Who Marked "Hate Much"	Item	Per Cent Who Marked "Hate Much"
To have bad habits	58	To make plans and drop *them*	34
To tell lies	55	When I do wrong	31
To bite fingernails	53	To forget things	28
To be bad	51	To sit in perfect silence	23
To be mean	51	A self-conscious feeling	21
To be unreliable	51	When I can't do as I like	19
To get in trouble	42	Always to want things	16
To be idle or lazy	37	To sit still	14
To do something I don't like	36		

The average distribution for all items, in terms of per cent, was: Like 11, Don't mind 12, Don't like 21, Hate 23, and Hate much 33.

hate or *hate much* by most of the girls. Also having bad habits, biting finger nails, or being lazy and forgetful receive the disapproval of many girls.

Sex Differences

A comparison of the sexes in relation to annoyances on 102 items listed by both groups indicates that girls are annoyed by more factors than boys and also that they react to a greater degree than the boys do to those annoyances. This is especially true in relation to personal conduct. Significant sex differences show greater dislike by boys of people who cheat. They are much more resentful than the girls of not being able to do as they like. In school, significantly more boys dislike arithmetic and singing, while more girls do not like to have nightwork to do. More boys do not like to sew or to hear their sisters sing. Significantly more girls hate to bite their finger nails.

Summary

A study of children's reactions to annoyances occurring in social relationships, home and family, school, and personal conduct indicates that many situations are extremely annoying to twelve-year-old children. Girls are more often and more extremely annoyed than boys.

Educational Implications. An annoyance is an unfavorable attitude which in many cases is caught from others. Some of the items listed should not be annoying to the normal child. People who are constantly irritated by every little thing reflect a failure to adjust to their environment. "They can't take it." On the other hand, a glance at many of the items listed in the tables suggests definite social standards which have been accepted by the children. Being annoyed by antisocial conduct reveals the acceptance of favorable group attitudes and leads to adaptation to the social group. But when children respond unfavorably to many unimportant situations, they should be taught to ignore those factors and not be disturbed by them. However, we should avoid as much as possible constant and unnecessary irritation. This is vital in a time when the human organism is being worn out by many irritations, fears, and worries.

Received January 20, 1944.

A School Survey of Eye-Hand Dominance

Gertrude Hildreth

Horace Mann-Lincoln School, Columbia University

The belief is widespread that mixed eye-hand dominance, that is, the tendency for eye and hand preference to be opposite sided, is a cause of reading disability, nervousness and behavior problems. The right-handed child who is left-eyed or vice versa is assumed to have more difficulty in learning to read and in making satisfactory personal adjustments. He is believed to mirror write more than the individual whose eye and hand dominance are like-sided. A parent recently inquired, "Isn't my child's reading difficulty due to the fact that tests show she is left-eyed?" The notion is also prevalent that mixed eye-hand dominance is a rare occurrence in the general population.

A number of published studies have given somewhat conflicting results. The present study was undertaken to obtain additional evidence on the subject, to determine the incidence of mixed dominance in an entire elementary school population and the association between mixed dominance and reading disability in the same population.

Results of Previous Studies

Incidence of Left-Eyedness and Mixed Dominance. Parson (8) reported 30 per cent left-eyedness, 12 per cent impartial, and 58 per cent right-eyedness in a normal school population in which left-handedness was present in about four per cent of the cases. Selzer (10) found 32 per cent of 96 children in the Cambridge schools, Grades II to VI, to be left-eyed; 72 per cent of the left-handed children were left-eyed. Schonell (9) investigated the relation between eye and hand dominance in 75 school children, with the following results (R.H. and R.E. signify right hand and right eye respectively): R.H.—R.E., 60%; R.H.—L.E., 25%; L.H.—L.E., 4%; L.H.—R.E., 3%; R.H.—and either eye, 8%; and L.H.—and either eye, 0.

Mixed Dominance and Reading Disability. Johnson O'Connor (7) reported an unusual amount of uncoordinated motion, twitching, shifting of position, speech irregularities and reading troubles among those who tested left-eyed and right-handed. He found less trouble among those whose handedness and eyedness agreed. A study by Mintz (5) indicated more mixed dominance among subnormal than among normal children. Monroe (6) reported that 43 per cent of the reading disability cases she studied showed mixed dominance. Fernald (1) reports on the problem as follows: "The right-handed cases and the cases of matched eye-hand dominance resemble the cases in which the dominance is not matched, are as serious in their deficiency, learn by the same methods, and are as successful in the final outcome.- - - The subject with unmatched eye and hand dominance learns to read and is able to read in an entirely normal manner with eye and hand dominance still opposite." Gates and Bond (2) found little association between eyedness, handedness and reading achievement. Johnson (4) reported no significant relationship between lateral dominance and reading disability. Wile (11) found among children showing reading disability, 62 per cent left-eyed, 8 per cent uncertain or mixed, 30 per cent right dominant. There is no indication as to how many of these pupils were left-handed. Witty and Kopel (12) found 33 per cent left-eyed in a group of poor readers, 31 per cent left-eyed in the normal reading, non-problem group. Wolfe (13) reported that more normal readers than retarded showed dominant left-eyedness on a sighting test and the groups were found not to be distinguished by handedness or a combination of eye-hand dominance. However, the retarded proved to be inferior in auditory functions, visual perception and emotional adjustment.

A School Survey

New evidence on the question was obtained through surveying an elementary school population consisting of 101 boys and 90 girls from kindergarten through Grade VI. The age range was from 6 to 11 with two children who lacked a month of being 6 and two who had just turned 12. The population rated above the average in general ability as measured by standard tests. Because of the selected nature of the population there were no cases in which reading disability could be attributed to lack of learning capacity.

Tests of Eye and Hand Dominance. Each pupil was given individual tests of eye and hand dominance. Four tests of eye dominance were used as follows:

1. *Sighting dot.* The child's attention was called to a black dot on a card tacked on the wall. He was told to stand with toes on a line marked on the floor three feet from the wall. The dot was placed at eye level. The child was given a box lid in the center of which was a half inch hole. He was asked to hold the hole at arms' length with both hands, he was told that one eye would be covered, then the other and that the dot would probably disappear when one eye was covered. The examiner covered each eye in turn, with a small card, asking the child to state whether he could or could not still see the dot. The eye with which he could still see the dot when one eye was covered was recorded as the dominant eye.

2. *Sighting with the Parson manoptoscope.* The child was asked to stand on a line two feet from the wall where the Parson board with red ball in between the letters L and R was hung up at the right height for his size. He was given the aluminum cone, was instructed to hold it in both hands, his attention was called to the red ball and he was instructed to fixate the red ball as he looked through the cone. Then the examiner explained that one letter could be seen alongside the ball, and the child was asked to state what letter

he could see. This letter was recorded. For the youngest children, small pictures of a dog and cat were used in place of the letters L and R.

3. *Sighting dot with the Parson cone.* The child standing two feet from a card with a black dot held by the examiner, was asked to fixate the dot through the aluminum cone. The position of the aperture of the cone with reference to an imaginary vertical line drawn through the child's nose was recorded. If to the right of the imaginary line, the child was recorded as right-eyed; if to the left, left-eyed.

4. *Peep Show.* The child was shown an imitation Easter egg with pictures inside that could be seen by looking through a quarter inch aperture at the smaller end of the egg. The examiner held the egg in both hands while the child studied and reported on the picture. The eye used in fixating the picture was recorded as the dominant eye.

A test was repeated if for any reason results were not clear cut. In the first three, the child kept both eyes open.

Only the fourth test in which the examiner rather than the child held the materials could be said to be entirely independent of handedness influence.

Determination of dominant eyedness was more difficult and probably less reliable in the case of younger children because of their difficulty in complying with instructions and their concern with irrelevant features of the test situation.

Tests of Handedness. Handedness was tested in three ways:

1. The child was requested to pick up a pair of scissors and to cut a corner from a piece of paper.
2. He was asked to write his name on the paper.
3. He was asked to pick up and toss a ball to the examiner several times.

These three tests afford a sampling of single handedness activities rather than a complete survey of each pupil's handedness habits.

Table 1

Results of Eye and Hand Dominance Tests in an Elementary School Population

	Eyedness					Handedness				
Age Level	4R	4L	3R	3L	2R–2L	3R	3L	2R	2L	Total Cases
5–6 Years										
Number	10	8	2	2	2	20	2	1	1	24
Per Cent	42	33	8.3	8.3	8.3	83.3	8.3	4.2	4.2	
7 Years										
Number	13	6	3	2	2	22	1	1	2	26
Per Cent	50	23	11.5	7.7	7.7	85	3.8	3.8	7.7	
8 Years										
Number	19	10	1	2	—	26	3	2	1	32
Per Cent	59	31	3.1	6.2	—	81	9.4	6.2	3.1	
9 Years										
Number	14	9	—	1	3	26	1	—	—	27
Per Cent	52	33⅓	—	3.7	11.1	96.3	3.7	—	—	
10 Years										
Number	27	8	4	2	1	34	4	—	4	42
Per Cent	64	19	9.5	4.8	2.4	81	9.5	—	9.5	
11 Years										
Number	12	15	3	5	5	36	2	1	1	40
Per Cent	30	37.5	7.5	12.5	12.5	90	5.0	2.5	2.5	

Indications of Reading Achievement. Objective test records, for the most part Stanford Achievement reading scores, or comparable data, were available for each pupil. Those who scored a year or more below grade median in the primary grades, a year and a half or more in the upper grades were considered disability cases for purposes of this study. Teachers' estimates and the necessity for recent remedial instruction in reading served as additional criteria in selecting the poorest readers in each class.

Results

Table 1 shows the number and percentage of pupils in each age level who were totally or partially right- or left-eyed, totally or partially right- or left-handed on the seven tests given. The right-eyed children

Table 2

Results of Eye and Hand Dominance Tests showing the Relationship between Eye and Hand Dominance

	Mixed Eye-Hand Dominance			
Age Level	R.H.-L.E.	L.H.-R.E.	R.H.-R.E.	L.H.-L.E.
5–6 Years				
Number	7	1	11	2
Total number mixed	8			
Total per cent mixed	33.3			
7 Years				
Number	8	3	14	—
Total number mixed	11			
Total per cent mixed	42			
8 Years				
Number	8	1	20	2
Total number mixed	9			
Total per cent mixed	28			
9 Years				
Number	9	—	14	1
Total number mixed	9			
Total per cent mixed	33.3			
10 Years				
Number	7	3	27	4
Total number mixed	10			
Total per cent mixed	24			
11 Years				
Number	20	3	12	—
Total number mixed	23			
Total per cent mixed	57.5			
Per cent mixed all age levels combined:	30.6			

The record for the only pair of twins:

	Eyedness Tests				*Handedness Tests*		
	1	2	3	4	1	2	3
Girl	R	R	R	R	R	R	R
Boy	L	L	L	L	R	R	R

were more consistently right-eyed in the four eyedness tests than were the left-eyed children. It was rather surprising to find a substantial number of the *right*-eyed children writing with the *left* hand.

There was no clear cut or consistent developmental tendency from one age group to the next. See Table 2. Although numerous research studies have shown a decrease in dominant left-handedness with age, and a slight tendency toward decreasing left-handedness was found in these results, no such decrease was shown in these data for eye-dominance.

Table 3

Eye-Hand Dominance Data for the Slowest Readers in Each Age Group

		Boys			Girls	
Age Level		Eyedness	Handedness		Eyedness	Handedness
7 years	Case 1	4R	3R		—	—
	Case 2	4L	3R		—	—
8 years	Case 1	4R	3L	Case 1	4L	3R (foreign child)
	Case 2	4R	3R	Case 2	4R	3R
				Case 3	4L	3R
9 years	Case 1	4L	3R	Case 1	4R	3R
	Case 2	4R	3R			
	Case 3	3L	3R			
10 years	Case 1	4L	3L	Case 1	4R	3R
	Case 2	4L	3R			
	Case 3	4R	3R			
	Case 4	4R	3R			
	Case 5	3L-1R	3R			
11 years	Case 1	3R-1L	3R	Case 1	2R-2L	3R
	Case 2	4R	3R			
	Case 3	4R	3R			
	Case 4	3R-1L	3L			

In fact, the eleven-year group showed an increase in left-eyedness over the ten-year group and more instability on the tests in general. The same situation did not prevail in the *handedness* test results for the eleven-year group. This may be a chance result due to the small population or it may be an indication of instability appearing with the approach of puberty.

Eye-Hand Dominance and Reading Disability. Twenty-two cases, sixteen boys and six girls in the entire group of 191, were selected as reading disability cases, according to the criteria given above. These pupils were all seven years of age or older. Table 3 shows the results for each age level in terms of eye-hand dominance.

Of the boys, there were seven or 44 per cent of the cases who showed mixed eye-hand dominance; of the girls, three or 50 per cent who showed this condition, a total of 45 per cent or less than half of all reading disability cases. The number of cases is too small to indicate reliably that these reading disability cases showed more tendency toward mixed dominance than the normally successful readers.

Although a somewhat larger proportion of mixed dominant cases had difficulty with reading than consistent dominant pupils, since fewer than half the mixed dominants were slow in learning to read, the conclusion must be drawn that mixed dominance is not a prevailing condition in reading disability, far less a dominant causal factor in the majority of disability cases.

Received December 10, 1943.

References

1. Fernald, G. R. *Remedial work in the basic skill subjects.* New York: McGraw-Hill, 1943.
2. Gates, A. I., and Bond, G. The relation of handedness, eyesighting and acuity dominance to reading. *J. educ. Psychol.*, 1936, 27, 450–456.
3. Hildreth, G. Bilateral manual performance, eye-dominance and reading achievement. *Child Develpm.*, 1940, 2, 311–317.
4. Johnson, P. W. The relation of certain anomalies of vision and lateral dominance to reading disability. *Monogr., Soc. Res. Child Develpm.*, 1942, 7, No. 2.
5. Mintz, A. A study of the indications of unstable unilateral cerebral dominance, reading disability and mental deficiency. Paper read at the Amer. Psychol. Ass. Meeting, Eastern Branch, 1943.
6. Monroe, M. *Children who cannot read.* Chicago: University of Chicago Press, 1932.
7. O'Connor, J. *Saturday Evening Post* article, February 27, 1943.
8. Parson, B. S. *Left-handedness, a new interpretation.* New York: Macmillan, 1924.
9. Schonell, F. J. *Backwardness in basic subjects.* London: Oliver and Boyd, 1942.
10. Selzer, C. A. *Lateral dominance and visual fusion.* Cambridge: Harvard University Press, 1933.
11. Wile, I. S. Eye dominance: its nature and treatment. *Arch. Ophthal., Chicago*, 1942, 28, 780–790.
12. Witty, P., and Kopel, D. Sinistral and mixed manual-ocular behavior in reading disability. *J. educ. Psychol.*, 1936, 27, 119–134. (Contains excellent summary and complete bibliography.)
13. Wolfe, L. S. Differential factors in specific reading disability. I. Laterality of function. *J. gen. Psychol.*, 1941, 58, 46–56; 57–70.

Book Reviews

Hahn, Eugene F. *Stuttering, significant theories and therapies.* Palo Alto: Stanford University Press, 1943. Pp. ix + 177. $2.00.

Professor Hahn has presented a compendium of theories and therapies of 25 of the "authorities in the field." He does not state the basis for his selection of specialists, although the list appears to be representative of the major educational and medical clinics doing research and therapeutic work on stuttering in this country and abroad, and to suggest the great variety of ways in which the problem is studied and treated. Since the several reports average less than six pages each, they are necessarily sketchy and incomplete. The author's purpose of facilitating comparisons of the philosophies and therapeutic procedures involved is inherently limited by this restriction. The careful student will want to go to original sources. The beginner may be more confused than enlightened by the multiplicity of points of view presented.

Each summary is divided into sections on *Theory* and *Therapy.* While most specialists appear to have been impelled to evolve a unique theory of stuttering, the careful reader will detect a considerable amount of similarity, under the verbiage, used to explain many of the theories. Some are certainly much more closely in line with modern psychological thinking than others. One of the values of the *theory* is no doubt the suggestion to the patient that the clinician is an expert. Anyone who has such a plausible explanation of the difficulty would surely appeal to the afflicted as a desirable person to administer treatment or direct his study. There is no indication that clinicians have consciously evolved their theories for this purpose. The multiplicity of theories is suggestive of the early reports to The French Academy of Sciences on the origin of language.

The Sections on *Therapy* appear to this reviewer as more significant. While few clinicians claim universal "cures" the success reported with such widely divergent treatments indicates either that stuttering is a malady of many causes and therefore subject to many treatments, or that there are common therapeutic values often overlooked in the various types of therapy, and not explained as such in the theory.

An Appendix of fifteen pages prepared by the author and entitled "Procedures in a Clinic for Stutterers" is a gesture toward unification of the various points-of-view of the "specialists." Within the limits of space used, this appears to be one of the better features of the book.

The book focuses attention on the great variety of backgrounds and approaches to the malady by those who are "authorities in the field," and leaves the reader with the realization that there is still much research to be done. If this book serves as a stimulus to further research and unification of thinking in the field, it will make an important contribution. This reviewer does not believe that the book will serve as such a stimulus.

Franklin H. Knower

The State University of Iowa,
Department of Speech

Cole, Luella. *Attaining maturity.* New York: Farrar and Rinehart, 1944. Pp. x + 212. $2.00.

This treatise is concerned with problems of adjustment in the modern world. Both in ordinary times and in times of stress there are always individuals who cannot adjust to complexities of life. Some of these never progress beyond childish attitudes and adolescent whims. Others regress to this status in trying times. In other words, such people never grow up, i.e., attain maturity. The mature individual is relatively free, he finds contentment and he feels secure. After noting the need for maturity, the author discusses the criteria of intellectual, emotional, social, and moral maturity. Popular escape is by fantasy, play, solitude, fanaticism, projection, sophistication and illness. Solutions for the mature person are briefly outlined. The final section is concerned with maturity and the war. The book contains many enlightening examples. Both the case studies and the discussion are well organized to aid in an analysis of acute problems of adjustment. As a guidebook for those seeking to attain maturity, however, the treatise is less adequate. There are a few ill-considered statements such as "the madhouse yawns for those who cannot be emotionally toughened" in war situations. Nevertheless the book has much to recommend it as "a guide to living with yourself and other people."

Miles A. Tinker

The University of Minnesota

Kingsley, Donald J. [Chairman]. *Recruiting applicants for the public service. A report submitted to the Civil Service Assembly by the Committee on Recruiting Applicants for the Public Service.* Chicago: Civil Service Assembly of the U. S. and Canada, 1942. Pp. xvi + 200. $3.00.

This volume sponsored by the Civil Service Assembly of the U. S. and Canada is another in the series dealing with the major phases of public personnel administration. The report is the work of J. Donald Kingsley aided by a committee consisting of Fred Zapolo, Russel Barthell, Irving Gold, George C. Brown, William Howell, George D. Halsey, and Edgar B. Young, among others.

The treatment of the subject matter can be subsumed under four categories or topics embracing six chapters: (1) historical antecedents of public service recruitment, (2) the nature of recruitment and recruitment procedures, (3) determining and forecasting personnel needs, and (4) application procedure and audit.

The content of the report is summarized as follows: Traditional concepts of recruitment are inadequate and have generally failed to produce a meritorious public service. The assumption that "keeping the rascals out"—a reaction to the spoils system—would of itself establish an adequate public service has been shown to be incorrect, and in fostering a laissez faire attitude in regard to recruitment, has resulted in a failure to attract men of superior ability to the public service. Recruitment should be positive in approach and should be to the lower levels of integrated classes of positions with promotion to higher levels. Such recruitment should occur at an early age and the point of entry into the service should be related to the educational system. The institution of these principles coupled with flexible transfer and promotion would give rise to a career system attractive to men of ability. Assumptions of such a career system are that adequate techniques for attracting and selecting recruits are available, and that high prestige value exists for the public service.

Recruitment is defined as "that process through which suitable candidates are induced to compete for appointments to the public service." The first step in the process is the determination of present and future personnel needs as a basis for planning the recruitment schedule. Such forecasting reduces the need for provisional appointments and serves to coordinate the recruitment, selection, and certification programs. Personnel needs may be determined from many sources but among the most useful are departmental demands for new personnel, personnel records and turnover statistics, analysis of departmental budgets, and surveys of employees for promotional possibilities.

There are two kinds of recruitment: anticipatory and direct. The former consists of building up favorable attitudes toward the public service without regard to any particular examination but with an eye to remote recruitment; the latter is the specific search and location of an adequate number of applicants for a specific examination.

Examination announcements are employed in direct recruitment to attract and interest qualified persons in examinations, to inform them of the nature and conditions of employment, and to discourage unqualified persons from applying. However, examination announcements are not equally suited for these functions and should only be employed for individuals already interested in public employment. Where legal requirements are not restrictive, examination publicity should be of a variegated nature and in line with principles of modern advertising. Announcements should be made more attractive and more care given to their distribution.

Admission to competition is invariably by application. This process involves the construction of an application form to secure the appropriate facts, administration of the application, determination whether minimum qualifications are met, and notification to the applicant of admission or rejection. The governing principle in application procedure is that only those items related to successful job performance, or statutory requirements, should be contained in the application, and audited in a manner most conducive to the determination of qualification or the lack of it.

Principal deficiencies of current recruitment procedures are: (1) unimaginative and stereotyped recruitment, (2) tendency to conform to legal minima, (3) absence of objective information upon which to base a sound selection program, and (4) a lack of planning and basic research.

The value of this report lies in its sound outline and treatment of the general problems of public service recruitment. It is not intended as a palliative indicating how more nurses, accountants, or social workers may be recruited in the war emergency but rather presents a number of principles upon which an adequate recruitment program should be based. This seems a wise departure for it is doubtful at best whether any volume of this scope can solve the problem of recruitment without a basic reformulation of governmental and administrative policies. Although the report is by no means novel in its recommendations, following closely the earlier studies of Professor Kingsley and the Commission of Inquiry on Public Service Personnel, its restatement is so fundamental as to make it mandatory reading. Nor does this detract from the excellent discussion of application procedure which comprises the second half of the book.

Of specific merit is the stress placed upon the need for research and objective information regarding basic recruitment and selection problems. The reiterated objections in the book to provisional or temporary appointments and the emphasis on forecasting of personnel needs as a corrective measure is wholesome. Of consequence also is the objection to long and involved application forms where the information requested has no relation to job success. Following industrial practice, applications should contain only those items that are of predictive significance in placement, in addition of course to those required by statute.

It is the reviewer's opinion that the career system as a cure-all for the recruitment problems of governmental agencies has been overstressed. Granting for the moment the assumptions of prestige for the public service and adequate selection and recruitment techniques—by no means safe assumptions—have not other factors been overlooked; e.g., competition of industry, business, or the professions; opportunities for higher salaries or profits which public jurisdictions could not hope to meet; the fact that an unknown quantity of individuals may not be interested in a life career in one agency or locality. Attraction of individuals to employment is always based upon a complex of factors of which the opportunity for a career is a variable weighted in varying degrees.

There is furthermore something to be said for recruitment of specialists at higher than the entrance level. Public jurisdictions are limited, and will remain limited by their very nature, in the scope and amount of training they can provide career neophytes in a large and varied number of occupations. Such training can usually be given only in professional and technical schools at considerable expense. Selection of numbers of recruits on the basis of appropriate aptitude may well be an insurmountable task for such a diversity of aptitude tests do not exist and public personnel aptitude testing is even less advanced than subject matter or achievement testing. Nor can a probationary period substitute for inadequate selection. The concept of a career system is laudable when due consideration is given related factors in recruitment.

Arthur Burton

California State Personnel Board

Journal of Applied Psychology

Vol. 29, No. 2 April, 1945

The Role of the Psychologist in Market and Advertising Research *

Wallace H. Wulfeck

The Federal Advertising Agency, Inc., New York City

Psychologists, generally, have been trained for the work of the laboratory, clinic and classroom. These places have been their occupational milieu. With the exception of the few, who in late years have been working in the field of business and industry, not many are acquainted with the actual details, the scope, or the wide variety of interests which make up the business psychologists' day. This paper will try to provide an insight into these processes. It will (*a*) take a quick glance at the rapid expansion of psychology's contribution to industrial and business development during the past ten years; (*b*) make mention of some of the personalities who have played a part; (*c*) give a job description by examples; along with (*d*) a prediction of the place of psychological research in the present and post-war economy.

Much of the discussion will be centered around the work of the Research Director in an advertising agency. This presents no particular limitation because most agencies serve a rather large list of accounts, representing a broad sample of manufacturers. In our case, it runs from corsets to railroads and touching such products as food, drug sundries, cosmetics, clothing, sanitary protection, pencils, magazines, baby food, heavy industry, and the emergency recruitment of farm and food processing labor. The agency research service reaches deeply into the areas of production, labor, and distribution of many of these clients as well as investigating their advertising problems. Consequently, agency research is in reality industrial and business research.

We will not become involved here with the broad speculative problems concerning the economic justification of advertising, or the kind of political-economic organization which should, or will, exist following this terrible catastrophe. We can only assume, as most politico-economic prognostications suggest, that there will be no fundamental change in the industrial organization of this country except perhaps for a continued exercise of limited governmental control over the practices of business and industry. One needs only to read Stuart Chase: *Where's the money coming from?*[1] to realize the kind of a world we will live in and the collossal obligation of a controlled capitalism to maintain economic security for "We The People." It is the continued existence of this kind of a world which provides the occupational challenge for psychologists now and tomorrow.

---

* Read at the February 1944 Meeting of the New York State Association for Applied Psychology.

The date at which applied psychology first gained a dawning recognition is debatable. For convenience, it may be placed around 1915–16 when Porter began publication of the *Journal of Applied Psychology*, or more specifically with the pioneering work of such men as Harlow Gale, Hugo Muensterberg, H. L. Hollingworth, E. K. Strong, and W. D. Scott. However, its great commercial impetus came much later in the business field following the developmental and promotional work of Link, Starch, and Gallup, among others. These men refined and expanded the now powerful methods known as marketing, public opinion and advertising measurement which have become useful tools in a wide field of socio-psychological, economic and military enterprises. Perhaps it is fair to say that by 1933 these methods and the national organizations to employ them effectively were reasonably well established.

The wide-spread acceptance of these and related methods has come only within the past ten years. In that brief span the growth has been phenomenal. The men and women recruited for this work came to it from many varied backgrounds and training. Almost none of them had training or experience in business, radio, or advertising. They brought to it one important element for success and that was a training grounded in scientific method, and a belief in its usefulness for solving the peculiar problems of business so far removed from the interests of the laboratory. Often they entered this field with grave misgivings, as though these fields of work were not a valuable or essential part of the life of our times, or would not allow uninfluenced freedom or scientific honesty in research. Notwithstanding, they came. Today the list of these people is formidable.

Who They Are

Here are a few of them taken from the membership lists of the A.P.A. and A.A.A.P. This list is by no means complete, and no one has been omitted with intent. A fairly large group have found their way into the services. There are two who have the M.A. degree, the rest hold the Ph.D.

---

[1] New York, 1943, The Twentieth Century Fund.

In Advertising Agency Research. George Gallup and Sam Hayes of Young and Rubicam, Otto Tinkelpaugh of J. M. Mathes Co., John B. Watson of Esty Co., Albert Blankenship of N. W. Ayer, Ernst Dichter of J. Sterling Getchell (now in radio), Alfred R. Root, and A. C. Welsh of Knox Reeves, Jean Pasmantier and W. H. Wulfeck of the Federal Advertising Agency.

With Independent Research Organizations. P. S. Achilles, Henry C. Link, Albert D. Freiberg, George Bennett, P. C. Corby, Eleanor Barnes of The Psychological Corporation; Daniel Starch of Daniel Starch Associates; Raymond Franzen and George Gallup of the Institute of Public Opinion.

Doing Radio Research. Frank Stanton, Ernst Dichter of the Columbia Broadcasting Company; Matthew Chappell with C. E. Hooper; Sidney Roslow, The Pulse of New York; Paul Lazarsfeld, and Robert H. McMurray.

In Government. Rensis Likert, and others.

Consultants. There is an important group of business, media, and radio research men holding faculty positions, among them D. B. Lucas and John Dollard, Readership; John Peatman, Radio; Hadley Cantril and Harry H. Field, Public Opinion; Arthur Kornhauser and Floyd Ruch, Marketing and Advertising; Harold E. Burtt and Howard P. Longstaff, Advertising.

It will be noted that two of those named are women recently come to this field. The work is wide-open to women having the interest, training, and personality for it.

Among business research leaders, psychologists tend to predominate as a professional group, but there are men from many other scientific disciplines holding prominent positions. They include economists, engineers, sociologists, and even chemists. Yes, psychology has made a place for itself in these fields even though it has failed to hold a dominant position in industrial and personnel work. Apropos of this last remark, it must be said in all fairness that the recently developed clinical approach to employment and personnel problems has stimulated new interest on the part of management, which may change the status of psychologists in industry.

If they are not already obvious, the reasons why psychologists enter the business research field are similar to those for doing any kind of clinical or experimental work.

1. As you will see, the problems are inherently interesting in themselves, as interesting as many reported in the current psychological literature.

2. Often the results have important economic, political or social value.

3. The rewards to the individual in satisfaction, recognition, and remuneration are equal to, if not greater than, those which accrue from work in other fields of psychological endeavor.

If you are to get a "feel" for the kinds of things these people do, the types of problems they attack, the scope of their operations, we can do no better than examine in outline a few of the projects studied by the research department of one agency during the past year. These problems are specific and representative and by no means exhaust the possibilities. Nor do they do more than indicate the ramifications of research in radio, marketing, and public opinion fields.

Examples of Jobs Done

First, it is well to know how the work is broken down into functional operations. There are three related types of investigation:

1. The application of psychological knowledge and research to the creative techniques in making effective advertising because the function of advertising is psychologic as well as economic. This includes evaluation and measurement of copy appeals, lay-out, art, and media. It tries to answer the question, "What is the most effective way to present the values in a given product so that the message will penetrate the minds of the consumer, become a part of his memory equipment, and result in persistent purchasing habits?"

2. The use of consumer survey methods to investigate product acceptance or rejection, penetration of advertising appeals, public attitudes toward business as a whole or some specific industry, opportunities for new markets, consumer buying power and preferences, readership of magazines and newspapers, and radio listening habits.

3. Library research—analysis of sales and distribution data, analysis of coverage and audiences of various media, study of the laws and regulations governing advertising, verification of claims made or to be made in ads, seeking and documenting incidents, events, personalities, and ideas to be used in advertisements. A multitude of other services, too numerous to mention, fall into this category of work.

Now to illustrate with specific examples the ways in which these types of research work out in fact. Projects "A" through "D" exemplify the first type of investigation.

Project A. You are aware of the existence of a group of magazines called the "Women's Service Books" written expressly to women. These magazines are advertised to advertisers just like any other product. Obviously the advantages in reaching a feminine audience with a product story in this type of publication, as against the general audience media, resides in the known fact that the interests, attitudes, and needs of women are different from those of men. In order to specify and dramatize these differences, an investigation was begun to canvass known differences in interests, abilities, attitudes, and achievement between the sexes. This study necessitated a search of the biological, sociological, anthropological, and psychological literature to establish a factual basis on which an advertising campaign could be built.

Project B. An investigation of the sexual attitudes, myths, prejudices, and taboos of women members of various religious faiths and subcultures which motivate against the use of tampons for menstrual protection—accompanied by a study of the types of advertising appeals most likely to combat these resistances.

Project C. The pre and post appearance evaluation of copy appeals through the use of consumer jury tests, consumer panels, split-runs, test city campaigns, coupon returns, and readership studies.

Project D. The use of Oleo-margarine has become a war-time necessity. In order to increase the use of this product, it was necessary to discover the housewife's attitudes, both favorable and unfavorable, toward its use and translate the information into an educational campaign to change her food use habits. Because of food scarcities and rationing this process has been carried on for many products, cereals for example.

Three kinds of investigation which characterize the second type of research follow:

Project E. One of the irksome problems of logistics was the protection of ordnance and machinery from corrosive moisture during storage and shipment. The method in general use was to immerse the materials in oil or grease. This was both time consuming and expensive. Scientists, put to work to develop a more practical device, discovered a compact crystalline chemical absorbent which could be packed along with the material in a sealed plio-film envelope. It was a simple, but complete solution.

The manufacturer of this product soon realized its value for dehumidifying purposes in homes and industry. It will be among the many new products available to make our living more comfortable in the post-war era. Our problem is to discover its potential uses in home and industry. At this moment hundreds of homes in the high humidity belt are experimenting with samples of the material and will tell us what it will do, how it can be used, how they like it, where it should be sold and for how much.

Project F. The research engineer of one of our clients was asked to develop a more accurate and sturdier device for the measurement of extreme temperatures in aircraft engines. Although the method cannot be revealed, he employed a simple electrical principle in making a new thermometer bulb which gives precision and hardiness characteristics far beyond the rigorous specifications set for it. The device may revolutionize remote temperature control. Our job is to investigate its possible uses in post-war industry, especially in the chemicals, plastics, and food processing industries, and to estimate its potential market.

Project G. Package design is an important element in the readiness with which women will buy a product. Given three different designs for a disposable tissue container submitted by a famous designer, which one will women be most likely to buy and why? Using a paired-comparison method, we asked a representative sample of women and got the answer.

As a sample of the third class of research here are two typical approaches.

Project H. War has brought many critical problems to the stocking industry. Problems of new synthetic yarns, new methods of manufacture, new styles, price control, and limitation on production. Mills have gone out of business and branded lines have lost position. These factors will have profound effects upon the structure of the business following the war. In order to prepare for any eventuality a client commissioned us to make a complete analysis of the industry covering factors effecting machinery, production, labor costs, sales, styles, and methods of distribution, pricing, merchandising, and advertising. This could all be done at a desk and is what we call library research. The resulting data, coupled with informed guesses based on experience, led to the formulation of several alternate plans or policies of operation which might be put into effect in the post-war era, depending upon the prevailing conditions.

Project I. No one had ever made a study of the economic importance and social significance of home-sewing to the life of the nation. The shortage of paper necessitated accurate information on which the government could base allocations of paper for the pattern industry, for patterns are the blue-prints, the basic tools, used in home-sewing. Exhaustive research into this problem revealed some interesting implications for the utilization of productive woman power, control of inflationary tendencies in the ready-to-wear industry, and the restraint of the cost of living for the great mass of wage earning families. We learned that more than 100 million garments were made at home during 1943, fifty-four per cent more than in 1939. That more than eighty-two per cent of all homes own sewing machines and that more than 500 million work hours were utilized in creative, useful labor, at an unbelievably small cost in paper tonnage. These facts are only a small part of the findings. The completed report makes a tenable case for home-sewing as an important influence in the control of a war economy.

You have had a rough picture of the scope and importance of the psychologists' work in business and industry. True, some of it is far afield from applied psychology, but none of it is remote from the application of scientific methodology or the study of aspects of human economic behavior. The logical next question to be answered here concerns the training which will best fit candidates for positions in this field.

Training for Business Research

In my opinion the ideal preparation will include scientific psychology, business and economic training, and practical experience to develop special skills, in that order of importance. The formal curriculum should include experimental and social psychology, scientific method, studies in motivation and behavior, and statistical methods. Other studies to round out the program should include sociology, economics, business administration, merchandising and marketing, advertising, and the techniques of marketing and public opinion research.

Finally a program of internship study of at least one year with an independent research organization such as the Psychological Corporation or the research department of a large advertising agency. With this background the beginner will be ready for one of the many junior positions on the road to a position of major responsibility.

Post-War Opportunities

Today there is a real occupational challenge for psychologists in business and industry. They have a genuine stake in the increased role research is to play in industry's program of production for use, and the vast problem of reconversion of industry. It seems clear that the major post-war problem facing this country will be the prevention of unemployment. The solution can lie only in increased production of consumer goods, for production means employment. Similarly, production requires purchasing power and markets which employment guarantees. The development of new markets, the lifting of the American standard of living so necessary for expanding employment opportunities will present increasingly difficult, widely ramified and interesting problems for research.

Business and industry is acutely aware of this condition and is expanding its research and planning facilities to cope with it. There is a rising tide of confidence in the specific contributions psychology can make to the solution of the many dilemmas facing them. During the past few years, independent research organizations have sprung up all over the country, some large industries have organized extensive departments of their own, while advertising agencies have expanded and strengthened their research programs to meet the needs of their clients. In the opinion of the leaders, this expansion will continue well into the period of economic readjustment.

Continued development will provide the opportunity, if the psychologists, who are best prepared to direct this work, are willing to demonstrate their interest and ability and to apply their science toward the solution of problems affecting the daily life of all of us.

Received March 29, 1944.

The Psychological Corporation Index of Public Opinion *

Henry C. Link

The Psychological Corporation, New York City

This is the eleventh in Series B of the nation-wide studies of public opinion conducted by The Psychological Corporation. Series A, generally known as the Psychological Barometer, was begun in September 1932, and is now the oldest periodic poll of public opinion and buying habits in existence. A total of 65 nation-wide surveys with 412,000 personal interviews has been made in this series which is currently being conducted quarterly with 10,000 home interviews each.

Series B was begun in September 1937 and several of the eleven surveys have been reported in this Journal. The present survey is based on 5,000 personal interviews made by 437 interviewers under the direction of 115 psychologists associated with The Psychological Corporation. The interviews were made between October 6 and 30, 1944, in 113 cities and towns representing a national cross-section of the urban population.

The exact questions used are given below, though not necessarily in their exact order. Most of the questions were asked of only half the sample, or in 2,500 interviews.

Wages and the Cost of Living

Q. "Is your family more prosperous (or better off) today than two years ago, less prosperous, or the same?"

Answers	In Oct. 1941	In Nov. 1942	In Oct. 1943	In Apr. 1944	In Oct. 1944
More prosperous	38%	29%	29%	28%	31%
About the same	47	47	46	48	46
Less prosperous	15	21	23	22	20
Don't know	—	3	2	2	3
Total interviews	2,000	2,500	2,500	2,500	2,500

In spite of the increases in the cost of living and heavier taxes, 77 per cent said they were as prosperous or more prosperous than they were two years ago. This increased or equal prosperity, as will be seen from the following table, is highest among the C and D, or the large factory and wage-earning groups. In these groups, constituting about 60 per cent of the urban population, over 80 per cent admit to unimpaired prosperity. But even in the predominantly white-collar and executive groups, A and B, over 73 per cent admit to continued prosperity.

---

* To conserve space and also paper further details of sampling and procedure are omitted in this report. For full details see H. C. Link, Tenth nation-wide experimental survey, *J. appl. Psychol.*, 1944, 28, 363–375. Also, to conserve space and paper only the barest explanatory comments are included with the present results. The full report may be obtained by writing to The Psychological Corporation, 522 Fifth Avenue, New York 18, New York.

		By Socio-Econ. Groups			
October 1944	Total	A	B	C	D
More prosperous	31%	28%	27%	34%	34%
About the same	46	46	46	46	47
Less prosperous	20	24	24	18	15
Don't know	3	2	3	2	4
Total interviews	2,500	250	750	1,000	500

Labor Riots vs. Race Riots

Q. "Which do you think will happen most often in the next two years, race riots or labor riots?"

		By Socio-Econ. Groups			
Answers	Total	A	B	C	D
Labor riots	45%	44%	44%	47%	41%
Race riots	31	31	38	30	24
Neither	7	9	5	8	9
Both	3	6	2	2	3
Don't know	14	10	11	13	23

The Roots of Free Enterprise and Post-War Planning

The individual American and his family constitute the roots of free enterprise, capitalism, and post-war planning. Post-war planning, to date, has been primarily a matter of government agencies on the one hand and business organizations and groups on the other. Therefore, it was decided to devote the larger part of the present survey to a study of the post-war plans being considered by individuals. However, instead of limiting the scope of this planning to articles they were going to buy, our inquiry included the kinds of jobs they were planning to hold or to obtain, plans for education and travel, and plans for going into business for themselves.

Q. "As you know, making jobs for the returning soldiers and many others is going to be a big problem. Have you heard or read about any business organizations who are planning for post-war jobs?"

54% answered "Yes."
46% answered "No."

Q. "Have you heard or read of any planning for post-war jobs by the Government or Government bureaus?"

62% answered "Yes."
38% answered "No."

Q. "Do you think that planning by the Government and business concerns will be enough, or do you think that the head of every family and every returning soldier should also make a post-war plan of his own?"

11% thought this would be enough.
80% thought individual plans also necessary.
9% were uncertain.

Q. "Have you or the head of your family made any post-war plans? That is, have you made any plans on how you are going to use your war-time savings and bonds when the war is over?"

Answers	Total	By Socio-Econ. Groups			
		A	B	C	D
Yes	54%	50%	58%	59%	44%
No	46	50	42	41	56

Q. "(If Yes in previous question) Would you mind telling me the things you and your family are going to *do* or *buy* in the order in which you are now planning to do them?"

Answers	Total
Home, farm, bungalow	22%
Household appliances	18
Automobile	17
Furniture and furnishings	9
Trip or vacation	7
Education	7
Household repairs	6
Go into business	4
Clothes	2
Miscellaneous	13

The emphasis in most studies in this field has been on buying rather than on doing and, as a result, distorted results have usually been obtained. The distortion is an overemphasis on the purchase of movable articles as compared with such items as education, traveling, going into business, building, redecorating. For most plans some capital and spending is necessary. However, our tests proved that the emphasis on spending tends to make people answer in the other extreme, which is saving, holding on to their bonds. For these reasons surveys previously made have produced conflicting reports. Some have shown primarily the enormous purchases which people were going to make. Others have indicated the large extent to which people were going to hold on to their savings.

Q. "Automobiles, washers, radios, houses, may be about 20 per cent higher in price when they come back. In that case would you still buy them?"

60% answered "Yes."
32% answered "No."
8% were uncertain.

Q. "(If answer was *No* in Q. 4) Why would you say you have made no plans?"

Answers	Total
Uncertain as to end of war	10%
Haven't enough bonds or savings	9
Haven't thought about it	5
Will hold bonds until maturity	4
Don't expect to need anything	4
No particular reason	3
Do not use savings to buy things	2
Total	40%

Q. "How much of the money you have saved since the war do you expect to use within a year or two after the war stops—all of it, two-thirds of it, one-third of it, or none of it?"

Of all respondents, 21 per cent said they were uncertain, while 7 per cent said they had no savings. The remainder were planning as follows:

48% planned not to spend any savings.
24% planned to spend one-third.
15% planned to spend two-thirds.
13% planned to spend all their savings.

Q. "Are you or the head of your family pretty sure of being able to keep your present job after the war?"

62% answered "Yes."
15% answered "No."
16% had their own business or profession.
7% were uncertain.

Q. "(If answer was *Yes*) Do you believe that your pay check will be as high, higher, or lower than now?"

28% answered as high.
15% answered lower.
6% answered higher.
13% were uncertain.

Q. "(If answer was *No*) If you think it likely that you will have to change from your present job into a peace-time job, are you fairly sure of getting such a job?"

8% answered "Yes."
6% answered "No."
1% were uncertain.

Q. "What are your difficulties in making a definite post-war plan for yourself or your family?"

No difficulties in planning	*37%*
Uncertainties of conditions after the war	*30*
I don't know what my job or income will be	7
I must wait for my husband or son to come home	0
Have no savings	4
Because of illness	3
Too old to plan; will retire soon	3
Miscellaneous reasons	*10*
Don't know; no answer	*7*

Q. a. "*If you had your choice (or: If you were choosing for a husband or a brother)* what work or occupation would you choose, after the war, of course?"

Q. b. "What are you doing now?"

Answers	Doing Now	Would Choose
Job with U. S. Government	7%	21%
Job with City or State Government	6	6
Job with large private company	25	25
Job with small private company	17	23
Own business	10	11
Professor, doctor, lawyer, teacher	7	6
In Army or Navy	9	*
Farmer	—	1
Miscellaneous, don't know, no man	10	6

* Less than .5 per cent

Received January 2, 1945.

A Method for Investigating Color Preferences in Fashions

B. R. Philip

Fordham University

Most investigations of fashion preferences have been based upon questionnaires. With the exception of Jacobson's (1) work little dimensional analysis under objective conditions has been attempted in this field. To determine whether the paucity of experimental work may be ascribable, in part at least, to the difficulty of adapting modern techniques to this field, the well-known psychophysical method of ranking was applied to the study of fashion preferences. Since the dimension selected was that of color, it was proposed incidentally, and by way of exemplifying the method, to determine whether men and women differ in their fashion preferences along that dimension. As this was an exploratory study only, few subjects were used, and the statistical techniques were carried out by the methods of small sampling.

One of the chief difficulties in an experimental study of fashion is the selection and presentation of the material. Obviously each subject must have ample opportunity to observe the material under the same standard conditions as every other subject. Were all observations made upon actual costumes, it is probable that other factors would tend to obscure the real facts, and validity would be secured at the expense of other qualities. Some procedure must be found to condense the material, to shorten the observation time and to equate observational conditions for all subjects.

Fortunately, instead of actual costumes, colored reproductions may be substituted with some confidence in the validity of the findings. People are accustomed to study fashions in this way, and often have recourse to pictures of garments to make their actual selections when purchasing clothes. One could proceed with much greater confidence if published experimental studies gave actual validity findings. To our knowledge, none have appeared in the literature. An unpublished study of the author, based on data from 19 men and 24 women in one group, and later, from 56 men and 106 women in another group, yielded average correlations of 0.85 for the men and 0.95 for the women, between preferences for actual material (Scotch plaid scarves), and colored pictures projected on a screen of these same scarves, when both rank order and paired comparison methods were employed. Accordingly it was decided to use colored fashion plates in lieu of actual costumes. In order to make sure that these costumes were in fashion, they were all selected from *Vogue*, for the current year, 1941, when this work was done.[1]

The material finally selected consisted of five sets of colored fashion plates. Each set consisted of 25 cards, all of the same size and showing one costume, commonly known as an afternoon dress; the costumes differed from card to card in some details, such as cut, ornamentation, etc. Within a set the costumes were of five different predominating colors, Black or Gray, Green, Blue, Red and Brown. There were five cards of each color, which differed considerably in hue, saturation and intensity; they were mostly monotone, although a few were striped or mottled. The sets differed among themselves in the height of the figure, which varied as follows: Set A, 11 inches; sets B and C, 9 inches; and sets D and E, 7 inches. The figures were cut out in silhouette and mounted individually on white cardboard, so as to leave an ample margin. Thus the size of the card in set A was 14½ × 8 inches, and in sets D and E, 8½ × 5 inches. A small code letter at the top of each card was inserted to permit of ready identification.

In the coding the sequence of colors was always the same, Black or Gray, Green, Blue, Red and Brown. Thus the letters, *a*, *f*, *k*, *p*, and *u* always designated a Black or Gray card, and *b*, *g*, *l*, *q*, and *v*, a Green card. This coding system was intended to permit easy segregation of the colors in tabulation. However in the actual experiment the cards were always presented in random order to the Ss, no one of whom noticed or understood the system used in coding.

The procedure was simple. College students, 19 men and 9 women, were given a set of cards randomly arranged to lay out, in order of preference, on a long table. They were provided with a set of typed instructions as follows:

> Rank the following set of cards according to preference, placing the card you most prefer in number one position, and placing the card you least prefer in the last (25th) position. Ratings are to be based upon general preference; all factors being taken into account. The question to be answered is simply, "Which outfit do you prefer?" No ties are allowed, but the cards may be shifted in position as often as preferred. Work quickly, and do not delay too long over any choice.

The sets were presented to the subjects in random order to equalize practice effects on the judgments of the various sets. All the rankings were entered on specially prepared tabulation sheets. One set only was ranked by every subject at the experimental sessions, which occurred bi-weekly. The purpose of spacing the experimental sessions was to minimize the "halo effect" and fatigue or boredom. To determine test-retest reliabilities, some six weeks after the ranking of the fifth set, the experiment was repeated for the women; the men were not available for this phase of the experimentation. By this procedure it was hoped that indications might be found whether or not there would be any time trend in the fashion preferences, while the costumes were still in fashion.

---

[1] Acknowledgment is here made to the Conde Nast Publications for supplying this material.

Results

The data were treated separately for men and for women. Linear scale values were computed by Guilford's method (2) assuming a composite standard based on the group of stimuli within the set, as a whole. Scale values were used rather than mean rankings in order to evaluate the dispersions of the judgments of one set as compared with another. The scale values in each set were then totalled for each group of colors, to form a five by five table of colors vs. sets, which was subjected to analysis of variance. As a check on the procedure, an analysis of variance using total ratings rather than scale values yielded substantially the same findings.

A point of some importance in this type of material is to determine how consistently the rankings are made by the individual and by the group. The degree to which the members of a group agree in their rankings may be inferred from the average intercorrelations of the rankings on any one set of cards. These intercorrelation values as determined by Kelly's formula, number 172, (3), and listed in Table 1, are positive and low, and average 0.220 for the women, and 0.134 for the men.

Table 1

Intercorrelations of Ratings from Different Sets of Cards

Set	Men	Women First Rating	Women Second Rating
A	0.169	0.203	0.153
B	0.135	0.267	0.175
C	0.135	0.327	0.327
D	0.095	0.171	0.204
E	0.136	0.137	0.243
Mean	0.134	0.221	0.220

The reliability of the individual rankings for the women, from a retest after six weeks, based on the correlations of the first and the second rankings of a set by the same individual, is relatively high, since an average reliability of 0.593 is found from all five sets (Table 2). Since the complete experiment is based on the rankings of five sets, were we

Table 2

Average Test-Retest Reliabilities for Women

Set	Per Subject	Per Card
A	0.644	0.801
B	0.608	0.835
C	0.553	0.885
D	0.504	0.700
E	0.560	0.813
Mean	0.503	0.825

to estimate the test reliability by applying the Spearman-Brown formula, the correlation becomes 0.879.

This is an average test-retest reliability *per subject;* but the average test-retest reliability may also be determined *per card.* It is obtained by correlating, by the rank-difference method, the total rating scores per card, yielded by the first and second ratings. This procedure is, in effect, a measure of the reliability of the group. These reliabilities are also presented in Table 2, from which it is seen that the average test-retest reliability *per card* is 0.825, which, when stepped up by the Spearman-Brown formula, becomes 0.959.

Accordingly the test employed is of satisfactory reliability, yielding fairly consistent results for the same individual, although individuals differ considerably among themselves in their preferences for any particular costume. Furthermore, after a lapse of six weeks, the group judgments are more consistent than the judgments of an individual. It is chiefly the reliability values found that give some confidence in the technique as here outlined for the investigation of fashions.

Since the experimental set-up was designed to investigate fashion preferences along a dimension of color by analysis of variance, it seemed of interest to do so, although it was recognized that the very small number of subjects would make any conclusions only very tentative and exploratory. The results are presented with this caution in mind, and are of course, only applicable to this group of subjects.

Analysis of variance discloses that the color of the costume is a statistically significant factor in the fashion preference for the men, but not for the women, in this experiment. For significance at the 1% and 5% level and with appropriate degrees of freedom, $F = 4.77$ and 3.01 respectively; the F values obtained were 18.63 and 2.26 for men and women respectively.

In order of preference the colors for the men were Blue, Black, Green, Red and Brown. The preferences for Blue were significantly greater at the 1% level than those for Green, Red and Brown; for Black, significantly greater at the 1% level than those for Red and Brown; and for Green, significantly greater than for Brown at the 5% level. As was stated, color preferences for the women are not statistically significant, but the order of preference was Green, Blue, Red, Brown and Black. Mean linear scale values, assuming a composite standard, are given in Table 3. This table implies that Blue costumes were on the average assigned by the men a scale position 0.259 standard deviation times higher than the mean rankings of all the cards in a set. Here the positive sign indicates a more favorable preference than the average.

Table 3

Mean Linear Scale Values for Colors

	Black	Green	Blue	Red	Brown
Men	0.132	−0.011	0.259	−0.161	−0.101
Women	−0.124	0.145	0.122	−0.060	−0.117

One might also determine the tendency to make preferred judgments for the colors by noting the percentage of times the various colors were given scale values greater than the mean scale value in each set. These percentages are given in Table 4, which confirms, in general, the findings based upon *mean scale values.*

Table 4

Percentages of Frequency of Preferred Selections for the Various Colors

	Black	Green	Blue	Red	Brown
Men	68	44	76	32	32
Women	48	68	56	48	36

The men are more prone to group around the mean judgments of costumes that have no marked positive or negative appeal; the women show greater dispersion of all judgments. Thus the relative proportions of judgments with scale values plus or minus 0.2 standard deviations from the mean, are 54% for the men, and 35% for the women; and for scale values plus or minus 0.4 standard deviations from the mean, 81% for the men and 67% for the women.

Since such a high proportion of cards were grouped about the mean, the implication is that no clear-cut preferences exist for many of the costumes, and definite preferences or rejections exist only for cards ranked in extreme positions. A listing of such preferences is given in Table 5, where the order of the color of the five most preferred costumes

Table 5

Distribution of First Five and Last Five Choices per Set According to Color

			First Choices			
	Black	Green	Blue	Red	Brown	Total
Men	8	2	11	3	1	25
Women, I	3	4	11	5	2	25
Women, II	4	4	11	4	2	25
			Last Choices			
	Black	*Green*	*Blue*	*Red*	*Brown*	*Total*
Men	2	—	1	10	12	25
Women, I	7	3	3	6	6	25
Women, II	7	2	3	7	6	25

The distribution of the choices for the first and for the second ratings are listed along the rows Women I and Women II.

per set corresponds quite well, for the men, but not for the women, with the linear scale values.

In the preferential positions the men in our experiment select Blue first and Black next, and in a decided manner they show least preferences for Brown and Red. The women also select Blue most frequently in the *preferential positions, with the other colors occurring in about equal* frequency. They are prone to spread their least preferred judgments among all the colors, with an indication of lesser allotment in those positions to Green and Blue. Green is the favorite choice for the women when all the cards are considered, as may be seen from the listing of the mean scale values, and from the percentages of preferred selections, although it does not receive as many preferred choices as does Blue. A point of interest is the close agreement of the first and second ratings of the women, confirming the reliability values found.

In summary we may state that this technique of investigating fashions is of adequate reliability. With the caution that this investigation has been carried out with few subjects, and that factors other than color such as style and pattern were not controlled, there are indications that the men place considerable emphasis on color in their rankings, Blue being the favorite color and Black the next. The women base their selections on a very composite criterion, as was pointed out by Hurlock (4) and by Barr (5), but Blue is for them also a favorite color, ranking slightly after Green. However their color preferences are subordinate to other factors in costume selection, and in this investigation were not statistically reliable. It is of interest to note that the colors selected for this investigation are the five most preferred colors as found by Barr (5) from her questionnaire, and the order of preference is closely the same in both investigations.

Received February 24, 1944.

References

1. E. W. Jacobson. An experimental investigation of the basic aesthetic factors in costume design. *Psychol. Monog.*, 1933. Vol. 18.
2. J. P. Guilford. *Psychometric methods.* N. Y.: McGraw-Hill, 1936, p. 200.
3. T. L. Kelly. *Statistical method.* N. Y.: Macmillan, 1924, p. 218.
4. E. B. Hurlock. Motivation in fashion. *Arch. Psychol.*, 1929, No. 111.
5. E. D. Barr. A psychological analysis of fashion motivation. *Arch. Psychol.*, 1943, No. 131.

Test Profiles as a Diagnostic Aid: The Minnesota Multiphasic Inventory

Lt Col Hermann O. Schmidt, AGD

Lincoln Army Air Field, Lincoln, Nebraska

The purpose of this paper is to present some findings in the use of the Minnesota Multiphasic Personality Inventory in a clinical situation.

No attempt will be made here to describe the development of the test, nor its applicability as a clinical instrument since these have been described elsewhere by its authors (1, 2, 3, 4, 5, 6, 7), by Leverenz (8), and by Schiele, Baker and Hathaway (9).

The test consists of 550 items printed upon cards. The subject is asked to file these cards into any one of three categories: *True*, *False*, *Cannot Say*. The test contains 12 scales: 3 validating scales, *Question* (cannot say), *Lie* and *False*; and 9 diagnostic categories: Hypochondriasis (H_s), Depression (D), Hysteria (H_y), Psychopathic Deviate (P_d), Masculinity-Femininity (M_f), Paranoia (P_a), Psychasthenia (P_t), Schizophrenia (S_c) and Hypomania (M_a).

The data have been gathered over a period of 9 months from cases studied in the Consultation Service of an Air Force replacement pool.

Procedure

The Multiphasic Inventory was administered as prescribed in the manual of directions (6). It was given in so far as practicable as part of the routine on the opening of a case. However, the pressure of time and the lack of sufficient test sets precluded the possibility of testing all cases coming to the Service.

All diagnoses were made by qualified neuropsychiatrists, employing classifications in common use in the Army, which in the main follow civilian practice.

This was not a controlled experiment, in the sense that careful selection of normal cases was sought, or that pure-culture clinical groups were (or could be) employed. Yet, there is reasonable certainty that the test did not influence the clinical diagnosis. Three of the four neuropsychiatrists making diagnoses avoided the test, merely finding it interesting *a postiori* that it seemed to agree with the clinical findings. The fourth neuropsychiatrist was quite interested in the test, but with a view to assisting in this validity study, he was careful not to see or to discuss the test on a particular individual until *after* he had made his clinical judgment.[1]

The Subjects

The subjects were all white male enlisted members of the Army Air Forces. The clinical groups had been referred to the Consultation Service because of difficulty in adjusting to the Army situation. The normal group was being re-evaluated, either because its records had become lost, were incomplete, or because there existed overages in particular occupational specialties, making reclassification necessary.

The deviate groups consisted of 121 subjects who fell into the diagnostic categories: constitutional psychopathic state, inadequate personality; constitutional psychopathic state, sexual psychopathy; psychoneurosis, mild; psychoneurosis, severe; and psychosis. There were several other groups, such as Organics, Simple Adult Maladjustment, Alcoholics, etc., but the cases were too few to have statistical significance and are not included here for that reason. Clinically, they were differentiable by profile. Table 1 gives descriptive data on the various groups as to number, age, intelligence, years of schooling, months of service, marital status, rank, number of states represented, population of the home town, and final disposition of the men who had been studied. This latter column consists of a percentage computation into 5 categories: discharge from the military service; limited military duty; full field duty; hospitalization (in most cases general hospital is implied) and duty (?). The last named category comprised those men on whom the follow-up data were incomplete, the only thing being known was that they were returned to duty, but whether on a full or limited duty status could not be ascertained.

The normal group N_o[2] consisted of 98 subjects, who upon inquiry manifested no evidence of overt or covert personality disorder. Their civilian and Army histories were devoid of any indication of disturbance: some had been overseas, but in a quiet theater; some had "washed-out" as aviation cadets, primarily for lack of psychomotor adaptability; but no trauma could be detected subjectively as attending either of these major sources of normal cases.

In all groups, by inspection, the descriptive data (Table 1) appeared to have a normal distribution generally and to be comparable. How-

---

[1] I should like to acknowledge here my indebtedness to 1st Lt. Roland A. Leslie, MC, for assistance and inspiration as a follow-worker and in the preparation of this paper.

[2] Hereafter the symbols N_o, C_p, C_H, P_m, P_s and P_{sy} will be used to designate the normal group; CPI, inadequate; CPI, sexual psychopathy; mild neurotics, severe neurotics and psychotic groups, respectively.

Table 1

Descriptive Data for the Normal Group (N_o) and the Constitutional Psychopathic State, Inadequate Personality (C_p), Constitutional Psychopathic State, Sexual Psychopathy (C_H), Psychoneurosis, Mild (P_m), Psychoneurosis, Severe (P_s), and Psychotic (P_{sy}) Groups

	Age (yrs.)			Education (yrs. comp.)			Service (months)			A.G.C.T.							Disch.
	$\bar{x}$	s	Range	$\bar{x}$	s	Range	$\bar{x}$	s	Range	$\bar{x}$	s	Range	Rank	Mar. Stat.	No. State Rep.	Pop. in 1000's	NFS FFD H Dy?
n, 98 N_o	24.9	3.5	18–36	11.7	1.6	8–17	38.3	27.8	10–204	114.3	12.0	78–139	M/Sgt. 9 T/Sgt. 24 S/Sgt. 16 Sgt. 17 Cpl. 11 Pfc. 5 Pvt. 16	M 50 S 47 D 1	34	0–1 18 1–10 22 10–100 33 100 25	100%
n, 11 C_p	24.4	3.6	20–31	11.0	2.7	6–15	22.8	27.9	4–108	105.7	17.5	75–128	Pvt. 8 Cpl. 2 Sgt. 1	S 8 M 3	6	0–1 1 1–10 1 10–100 2 100 7	27.3 9.1 — 9.1 54.5
n, 7 C_H	27.3	6.4	20–40	12.9	3.4	8–18	11.1	8.3	4–30	105.2	16.1	74–121	Pvt. 3 Pfc. 1 Cpl. 1 Sgt. 2	S 6 M 1	5	0–1 0 1–10 0 10–100 1 100 6	85.7% — — — 12.3%
n, 26 P_m	24.6	4.5	19–37	11.5	2.3	8–19	12.7	10.6	4–54	117.6	15.7	76–142	Pvt. 6 Pfc. 9 Cpl. 5 Sgt. 4 S/Sgt. 1 T/Sgt. 1	S 13 M 13	16	0–1 4 1–10 3 10–100 6 100 13	7.7 19.3 3.8 — 69.2
n, 64 P_s	24.7	4.5	19–36	11.3	1.8	3–19	13.9	10.5	2–42	108.4	21.3	74–145	Pvt. 27 Pfc. 16 Cpl. 8 Sgt. 8 S/Sgt. 3 T/Sgt. 2	S 34 M 30	27	0–1 6 1–10 8 10–100 11 100 39	50.0 21.9 9.4 4.7 14.0
n, 13 P_{sy}	24.9	5.5	19–34	10.4	1.9	6–13	26.7	38.7	3–144	108.2	11.9	86–129	Pvt. 5 Pfc. 6 T/Sgt. 2	S 8 M 4 D 1	10	0–1 1 1–10 2 10–100 5 100 5	53.8 7.7 15.4 7.7 15.4

ever, the intellectual distribution is skewed somewhat to the left, in that no Army General Classification Test group V's were given the test. This was because of the highly questionable ability of this low group to comprehend the various test items.

Results

Figures 1–5 show graphically the T-score profiles of the C_p, C_H, P_m, P_s and P_{sy} clinical groups together with the N_o group.

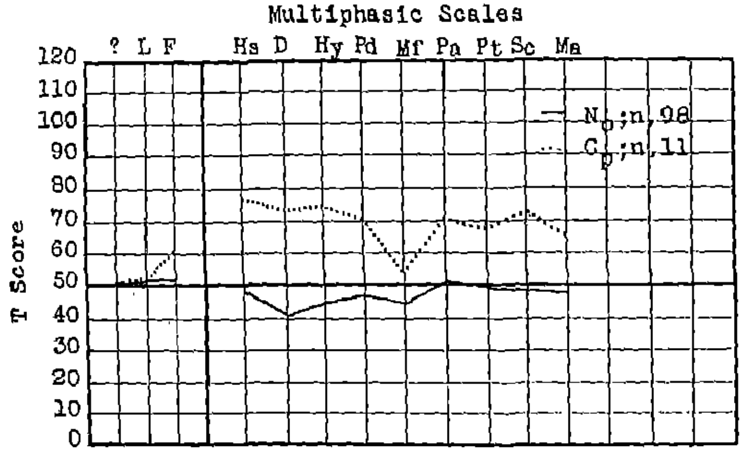


Fig. 1. T-score profiles on the Multiphasic Inventory for the Normal (N_o) and constitutional psychopathic state, inadequate personality (C_p) groups.

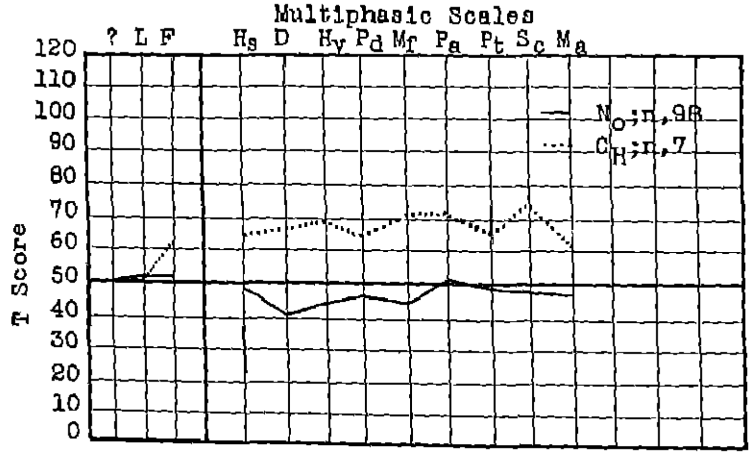


Fig. 2. T-score profiles on the Multiphasic Inventory for the Normal (N_o) and constitutional psychopathic state, sexual psychopathy (C_H) groups.

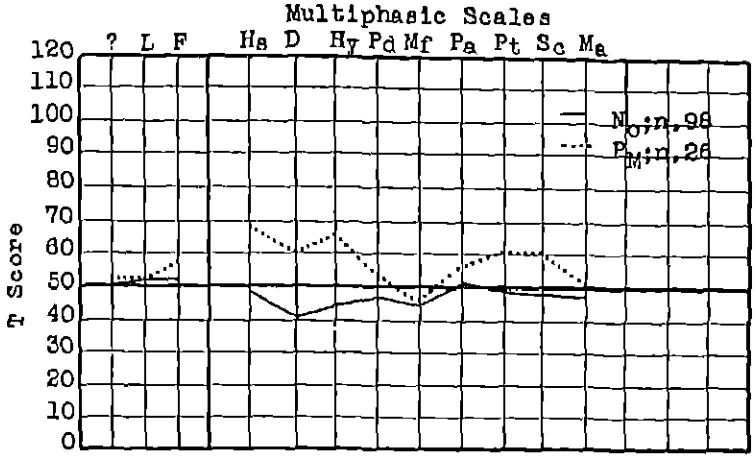


FIG. 3. T-score profiles on the Multiphasic Inventory for the Normal (N_o) and psychoneurosis, mild (P_m) groups.

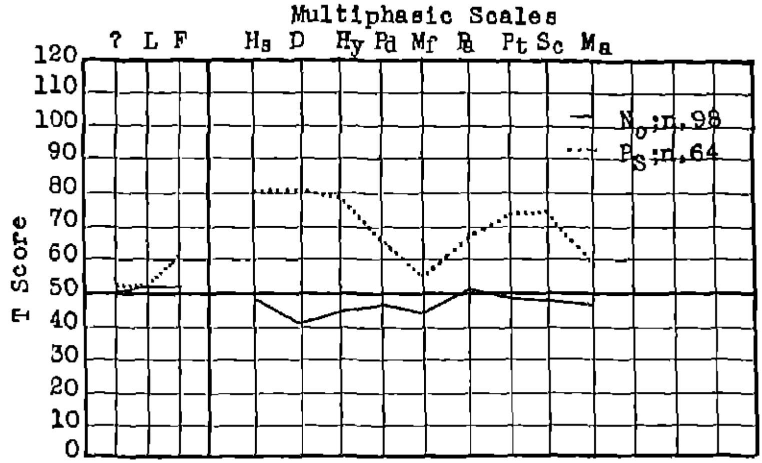


FIG. 4. T-score profiles on the Multiphasic Inventory for the Normal (N_o) and *psychoneurosis, severe* (P_s) *groups.*

Table 2 presents the means, sigmas and ranges of all groups; and Tables 3, 4, 5, 6 and 7 give the differences, standard errors of difference and critical ratios between the various groups (T-score values, uncorrected for small samples or beyond 1 decimal point).

It is at once apparent from inspection of Figures 1–5 that the deviate groups present different and characteristic profiles from those made by the normal group.

The profile for the N_o group is reasonably flat and approaches closely the T-score mean level of 50. It is somewhat lower generally than this mean range of 50 and lower than the normal curve found by Leverenz (8). Within its own configuration the low point is reached on the depression scale; while the high point is for paranoia.

The profiles for the deviate groups show obvious divergence from the normal profile. In all instances the curves are higher, being less so for the P_m group and maximum for P_{sy}. The depression scale appears as an index to the seriousness of the difficulty. It is higher than either hypochondriasis or hysteria in the P_s and P_{sy} groups, and lower for all the others. In the case of the C_H group, the profile rises on the masculinity-femininity scale whereas in all the other groups the curve is depressed at this point. By way of indicating characterizing features of the various profiles, it may be noted that:

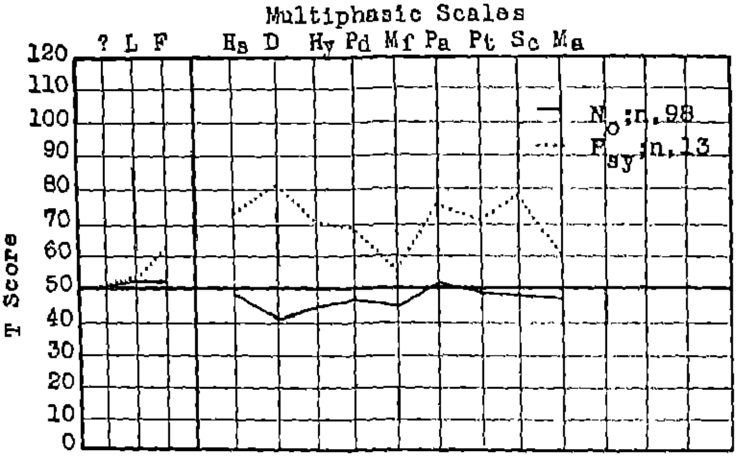


FIG. 5. T-score profiles on the Multiphasic Inventory for the Normal (N_o) and psychosis (P_{sy}) groups.

(*a*) The N_o profile, although generally a straight line at T-score mean level, shows a low-point at depression and a high-point at paranoia;

(*b*) The C_p curve has high points about 2 sigmas above mean level through the psychopathic deviate scale, drops to nearly normal at masculinity-femininity, rises to paranoia, drops again slightly on psychasthenia, rises to schizophrenia and falls off on hypomania;

(*c*) The C_H profile is approximately a straight line, between 1 and 2 sigmas above the mean level, with high-points on the masculinity-femininity and schizophrenia scales;

(*d*) The P_m profile, at about 1 sigma above normal, descriptively shows a high level at hypochondriasis-depression-hysteria (depression lowest), dropping sharply on masculinity-femininity and rising sharply again to psychasthenia-schizophrenia;

Table 2

T-Score Means, Standard Deviations and Ranges for the Normal Group (N_o), Constitutional Psychopathic State, Inadequate Personality (C_p), Constitutional Psychopathic State, Sexual Psychopathy (C_H), Psychoneurosis, Mild (P_m), Psychoneurosis, Severe (P_s), and Psychotic (P_{sy}) Groups, on the Separate Scales of the Minnesota Multiphasic Inventory

	N_o n, 98			C_p n, 11			C_H n, 7			P_m n, 26			P_s n, 64			P_{sy} n, 13		
	$\bar{x}$	s	Range	$\bar{x}$	s	Range	$\bar{x}$	s	Range	$\bar{x}$	s	Range	$\bar{x}$	s	Range	$\bar{x}$	s	Range
?	50.1	.5	50–53	51.7	4.1	50–64	50.0	0.0	—	51.9	5.9	50–80	51.1	4.6	50–80	50.8	2.7	50–60
L	52.1	3.5	50–63	52.8	3.1	50–60	50.0	0.0	—	51.7	4.1	50–70	51.9	4.0	50–66	52.5	6.9	50–76
F	52.2	4.1	50–68	61.0	9.1	50–80	61.7	8.1	50–80	56.9	8.1	50–80	60.9	9.9	50–80	63.0	10.8	50–80
H_s	48.5	8.2	40–90	77.1	18.6	47–117	65.7	11.4	42–85	69.9	16.0	47–99	80.3	15.9	40–108	74.6	16.2	53–106
D	41.2	12.7	29–96	73.7	19.3	41–99	66.6	12.4	29–82	60.6	21.4	29–94	80.7	20.8	39–120	82.1	27.5	32–116
H_y	45.1	10.0	35–89	75.1	20.4	40–100	69.0	12.9	47–93	66.6	19.7	35–106	79.3	15.6	49–109	70.6	20.0	35–111
P_d	47.0	13.1	25–86	71.0	17.9	30–96	66.4	16.1	35–86	53.2	14.9	32–101	66.3	17.1	32–106	69.4	23.2	20–111
M_f	45.2	9.7	26–73	54.4	10.7	37–76	71.4	17.2	44–94	46.9	11.5	28–74	56.5	11.1	35–102	55.5	13.9	37–78
P_a	51.0	8.9	33–79	71.5	14.7	47–102	71.4	8.4	44–85	56.0	14.4	33–85	66.2	11.6	38–114	76.2	25.1	38–117
P_t	49.5	8.4	38–78	68.0	14.4	45–89	66.3	9.5	45–77	60.5	14.9	39–93	72.9	12.7	45–95	71.9	15.7	42–98
S_c	48.5	7.7	38–71	72.6	19.9	45–105	74.7	12.6	45–90	60.7	15.1	40–103	73.6	15.9	43–106	78.5	18.6	45–114
M_a	48.2	9.4	30–75	66.6	12.4	43–86	61.0	8.9	45–70	50.4	12.1	30–72	59.5	11.5	34–88	61.5	13.5	30–84

Table 3

mparison of the Normal Group (N_o) and the Clinical Groups of Constitutional Psychopathic State, Inadequate Personality (C_p), Constitutional Psychopathic State, Sexual Psychopathy (C_H), Psychoneurosis, Mild (P_m), Psychoneurosis, Severe (P_s), and Psychosis (P_{py}) on the Separate Multiphasic Inventory Scales Showing Difference, Standard Error of Difference and Critical Ratio of T-scores*

n, 98 P_m n, 26 n, 11 P_s n, 64 n, 7 P_{py} n, 13	N_o–C_p			N_o–C_H			N_o–P_m			N_o–P_s			N_o–P_{py}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	1.6	1.2	1.3	– .1	.05	2.0	.3	1.1	.7	1.1	.6	.2	.7	.7	1.0
L	.7	1.0	.7	–2.1	.3	7.0	–.4	.9	.4	–.2	.6	.3	.4	1.6	.3
F	8.8	2.8	3.1	9.5	3.7	2.6	4.7	1.6	2.9	8.7	1.3	6.7	10.8	3.0	3.6
H_s	28.6	5.7	5.0	17.2	4.4	3.9	21.4	3.2	6.7	31.8	2.1	15.1	26.1	4.5	5.8
D	32.5	5.9	5.5	25.4	4.9	5.2	19.4	4.4	4.4	39.5	2.9	13.6	40.9	7.8	5.2
H_y	30.0	6.2	4.8	23.9	4.9	4.9	21.5	3.9	5.5	34.2	2.2	15.5	25.5	5.6	4.6
P_d	24.0	5.5	4.4	19.4	6.2	3.1	6.2	3.2	1.9	19.3	2.5	7.7	22.4	6.6	3.4
M_f	9.2	3.3	2.8	26.2	6.6	4.0	1.7	2.5	.7	11.3	1.7	6.6	10.3	3.9	2.6
P_a	20.5	4.5	4.6	20.4	3.3	6.2	5.0	2.9	1.7	15.3	1.7	9.0	25.2	7.0	3.6
P_t	18.5	4.4	4.2	17.8	3.7	4.8	11.0	3.0	3.7	23.4	1.8	13.0	22.4	4.4	5.1
S_c	24.1	6.1	3.9	26.2	4.8	5.5	12.2	3.1	3.9	25.1	2.1	12.0	30.0	5.2	5.8
M_a	18.4	3.8	4.8	12.8	3.5	3.7	2.2	2.5	.9	11.3	1.7	6.6	13.3	3.8	3.5

Difference is first of a pair from second; hence, minus signs may occur.

Table 4

Comparison of the Constitutional Psychopathic State, Inadequate Personality Group (C_p) and the Clinical Groups of Constitutional Psychopathic State, Sexual Psychopathy (C_H), Psychoneurosis, Mild (P_m), Psychoneurosis, Severe (P_s), and Psychosis (P_{sy}) on the Separate Multiphasic Inventory Scales Showing Difference, Standard Error of Difference and Critical Ratio of T-scores*

C_p n, 11 P_s n, 64 C_H n, 7 P_{sy} n, 13 P_m n, 26	C_p–C_H			C_p–P_m			C_p–P_s			C_p–P_{sy}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	D/SE_{dif}	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	−1.7	1.2	1.4	.2	1.7	.1	− .6	1.3	.5	− .9	1.5	.6
L	−2.8	.9	3.1	−1.1	1.2	.9	− .9	1.1	.8	− .3	2.1	.1
F	.7	4.1	.2	−4.1	3.2	1.3	.8	3.0	.3	2.0	4.1	.5
H_s	−11.4	7.1	1.6	−7.2	6.4	1.1	2.2	5.9	.4	−2.5	7.2	.3
D	−7.1	7.4	1.0	−13.1	7.2	1.8	7.0	6.3	1.1	8.4	9.6	.9
H_y	−6.1	7.7	.8	−8.5	7.2	1.2	4.2	6.4	.7	−4.5	8.2	.5
P_d	−4.6	8.1	.6	−17.8	6.1	2.9	−4.7	5.8	.8	−1.6	8.4	.2
M_f	17.0	7.2	2.4	−7.5	3.9	1.9	2.1	3.5	.6	1.1	5.0	.2
P_a	− .4	5.4	.1	−15.5	5.2	2.9	−5.2	4.6	1.1	4.7	8.2	.6
P_t	−1.7	5.6	.3	−7.5	5.2	1.4	4.9	4.6	1.1	3.9	5.9	.7
S_c	2.1	7.7	.3	−11.9	6.7	1.8	1.0	6.3	.2	5.9	7.9	.7
M_a	−5.6	5.0	1.1	−16.2	4.4	3.7	−7.1	4.0	1.8	−5.1	5.0	1.0

* Difference is first of a pair from second; hence, minus signs may occur.

Table 5

Comparison of Constitutional Psychopathic State, Sexual Psychopathy (C_H), and the Clinical Groups of Psychoneurosis, Mild (P_m), Psychoneurosis, Severe (P_s), and Psychosis (P_{sy}) on the Separate Multiphasic Inventory Scales Showing Difference, Standard Error of Difference and Critical Ratio of T-scores*

C_H n, 7 P_s n, 64 P_m n, 26 P_{sy} n, 13	C_H–P_m			C_H–P_s			C_H–P_{sy}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	1.9	1.1	1.7	1.1	.5	2.2	.8	.8	1.0
L	1.7	.8	2.1	1.9	.5	3.8	2.5	1.6	1.6
F	−4.8	3.4	1.4	−.8	3.3	.2	1.3	4.2	.3
H_s	4.2	5.3	.8	14.6	4.7	3.1	8.9	6.2	1.4
D	−6.0	6.3	1.0	14.1	5.4	2.6	15.5	8.9	1.7
H_y	−2.4	6.2	.5	10.3	5.3	1.9	1.6	7.3	.2
P_d	−13.2	6.8	1.9	−.1	6.4	.2	3.0	8.9	.8
M_f	−24.5	6.9	3.6	−14.9	6.7	2.2	−15.9	7.6	2.1
P_a	−15.4	4.2	3.7	−5.1	3.5	1.5	4.8	7.6	.6
P_t	−5.8	4.6	1.3	6.6	3.9	1.7	4.6	5.6	.8
S_c	−14.0	5.6	2.5	−1.1	5.1	.2	3.8	7.0	.5
M_a	−10.6	4.1	2.6	−1.5	3.8	.4	.5	5.0	.1

* Difference is first of a pair from second; hence, minus signs may occur.

Table 6

Comparison of psychoneurosis, mild (P_m), and the clinical groups of psychoneurosis, severe (P_s), and psychosis (P_{sy}) on the separate Multiphasic Inventory scales showing difference, standard error of difference and critical ratio of T-scores*

P_m n, 26 P_s n, 64 P_{sy} n, 13	P_m–P_s			P_m–P_{sy}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	−.8	1.3	.6	−1.1	1.4	.8
L	.2	.9	.2	.8	2.1	.4
F	4.0	2.0	2.0	6.1	3.4	1.8
H_s	11.4	3.7	3.1	4.7	5.5	.9
D	20.1	4.9	4.1	21.5	8.7	2.5
H_y	12.7	4.3	3.0	4.0	6.8	.6
P_d	13.1	3.6	3.6	16.2	7.1	2.3
M_f	9.6	2.7	3.6	8.6	4.4	1.9
P_a	10.3	3.3	3.1	20.2	7.5	2.7
P_t	12.4	3.3	3.8	10.4	5.2	2.0
S_c	12.9	3.6	3.6	17.8	5.9	3.0
M_a	9.1	2.8	3.3	11.1	4.4	2.5

* Difference is first of a pair from second; hence, minus signs may occur.

Table 7

Comparison of psychoneurosis, severe (P_s), and the clinical group of psychosis (P_{sy}) on the separate Multiphasic Inventory scales showing difference, standard error of difference and critical ratio of T-scores*

P_s n, 64 P_{sy} n, 13	P_s–P_{sy}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	− .3	.9	.3
L	.6	1.9	.3
F	2.1	3.2	.7
H_s	−5.7	4.9	1.2
D	1.4	8.1	.2
H_y	−8.7	5.9	1.5
P_d	3.1	6.8	.5
M_f	−1.0	4.4	.2
P_a	9.9	7.1	1.4
P_t	−1.0	4.6	.2
S_c	4.9	5.5	.9
M_a	2.0	4.0	.5

* Difference is first of a pair from second; hence, minus signs may occur.

Table 5

Comparison of Constitutional Psychopathic State, Sexual Psychopathy (C_H), and the Clinical Groups of Psychoneurosis, Mild (P_m), Psychoneurosis, Severe (P_s), and Psychosis (P_{sy}) on the Separate Multiphasic Inventory Scales Showing Difference, Standard Error of Difference and Critical Ratio of T-scores*

C_H n, 7 P_s n, 64 P_m n, 26 P_{sy} n, 13	C_H–P_m			C_H–P_s			C_H–P_{sy}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	1.9	1.1	1.7	1.1	.5	2.2	.8	.8	1.0
L	1.7	.8	2.1	1.9	.5	3.8	2.5	1.6	1.6
F	−4.8	3.4	1.4	− .8	3.3	.2	1.3	4.2	.3
H_s	4.2	5.3	.8	14.6	4.7	3.1	8.9	6.2	1.4
D	−6.0	6.3	1.0	14.1	5.4	2.6	15.5	8.9	1.7
H_y	−2.4	6.2	.5	10.3	5.3	1.9	1.6	7.3	.2
P_d	−13.2	6.8	1.9	− .1	6.4	.2	3.0	8.9	.8
M_f	−24.5	6.9	3.6	−14.9	6.7	2.2	−15.9	7.6	2.1
P_a	−15.4	4.2	3.7	−5.1	3.5	1.5	4.8	7.6	.6
P_t	−5.8	4.6	1.3	6.6	3.9	1.7	4.6	5.6	.8
S_c	−14.0	5.6	2.5	−1.1	5.1	.2	3.8	7.0	.5
M_a	−10.6	4.1	2.6	−1.5	3.8	.4	.5	5.0	.1

---

* Difference is first of a pair from second; hence, minus signs may occur.

Table 6

Comparison of psychoneurosis, mild (P_m), and the clinical groups of psychoneurosis, severe (P_s), and psychosis (P_{sy}) on the separate Multiphasic Inventory scales showing difference, standard error of difference and critical ratio of T-scores*

P_m n, 26 P_s n, 64 P_{sy} n, 13	P_m–P_s			P_m–P_{sy}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	−.8	1.3	.6	−1.1	1.4	.8
L	.2	.9	.2	.8	2.1	.4
F	4.0	2.0	2.0	6.1	3.4	1.8
H_s	11.4	3.7	3.1	4.7	5.5	.9
D	20.1	4.9	4.1	21.5	8.7	2.5
H_y	12.7	4.3	3.0	4.0	6.8	.6
P_d	13.1	3.6	3.6	16.2	7.1	2.3
M_f	0.6	2.7	3.6	8.6	4.4	1.9
P_a	10.3	3.3	3.1	20.2	7.5	2.7
P_t	12.4	3.3	3.8	10.4	5.2	2.0
S_c	12.0	3.0	3.6	17.8	5.0	3.0
M_a	0.1	2.8	3.3	11.1	4.4	2.5

* Difference is first of a pair from second; hence, minus signs may occur.

Table 7

Comparison of psychoneurosis, severe (P_s), and the clinical group of psychosis (P_{sy}) on the separate Multiphasic Inventory scales showing difference, standard error of difference and critical ratio of T-scores*

P_s n, 64 P_{sy} n, 13	P_s–P_{sy}		
	Dif.	$SE_{dif.}$	$D/SE_{dif.}$
?	− .3	.9	.3
L	.0	1.9	.3
F	2.1	3.2	.7
H_s	−5.7	4.9	1.2
D	1.4	8.1	.2
H_y	−8.7	5.0	1.5
P_d	3.1	6.8	.5
M_f	−1.0	4.4	.2
P_a	9.9	7.1	1.4
P_t	−1.0	4.6	.2
S_c	4.9	5.5	.9
M_a	2.0	4.0	.5

* Difference is first of a pair from second; hence, minus signs may occur.

(*e*) The P_s curve shows a high plateau, 3 sigmas above normal at hypochondriasis-depression-hysteria (depression slightly higher), drops at masculinity-femininity, and rises again to schizophrenia;

(*f*) The P_{sy} profile shows a high peak on depression at 3 sigmas above normal, a descending to masculinity-femininity to nearly normal, a sharp rise (2.5 sigmas) to paranoia, a slight drop at psychasthenia and a rise to nearly 3 sigmas at schizophrenia.

There is a decrease on all profiles on the hypomania scale.

Of the validating indicators, ?, L and F, the curve rises more sharply at F for the P_{sy} group and has minimal accentuation for N_o.

From examination of Table 3, where the N_o group is compared with the clinical groups, C_p, C_H, P_m, P_s and P_{sy}, it will be noted that in all but 6 instances a significant difference is apparent, the critical ratio being 3.1 or higher—ranging from 3.1 for N_o–C_H on the P_d scale to 15.5 for N_o–P_s on the H_y scale.[3] Of the 6 cases that diverge, 4 are in the N_o–P_m comparison on the psychopathic deviate, masculinity-femininity, paranoid and psychasthenic scales, with critical ratios of 1.9, .7, 1.7 and .9 respectively. The remaining 2 are on the masculinity-femininity scale for N_o–C_p and N_o–P_{sy}, with critical ratios of 2.8 and 2.6 respectively.

In the matter of the validating items of ?, L and F scales, no significant differences appear on the ? scale, although between N_o–C_H the critical ratio is 2.0, the N_o group having the higher ? score. On the L scale, the only difference possessing significance is that between N_o–C_H, where the critical ratio is 7.0, and where again the N_o group makes the higher score. The F scale shows significant differences in 3 of the comparisons, and borderline significance in the remaining 2. In each instance the greater F score is made by the clinical group. The critical ratios in ascending order are respectively: N_o–C_H, 2.6; N_o–C_p, 3.1; N_o–P_{sy}, 3.6, and N_o–P_s, 6.7.

Table 4 compares the C_p group with the other clinical groups. In only 2 instances here are there significant differences: the L scale has a critical ratio of 3.1 for C_p–C_H, the C_p score being the higher; and the critical ratio for C_p–P_m on hypomania is 3.7, the C_p score being higher. Approaching significance is the critical ratio of 2.4 for C_p–C_H on the masculinity-femininity scale, the C_H group being the higher; C_p–P_m on the psychopathic deviate scale shows a critical ratio of 2.9, the C_p group being higher; and on the paranoid scale, C_p–P_m, the critical ratio is 2.9, C_p being the higher.

Table 5 compares the C_H group with the clinical groups P_m, P_s and P_{sy}. In the validating scales, only the L score for C_H–P_s shows a difference that is significant—the critical ratio is 3.8, with C_H being higher.

---

[3] In this and succeeding discussion of significance, a critical ratio of 3.0 is taken as the criterion for significance.

On the masculinity-femininity and paranoia scales for C_H–P_m, and hypochondriasis on C_H–P_s the critical ratios are respectively 3.6, 3.7 and 3.1, with C_H higher in the first 2 instances, and the lower score in the latter. These are the only significant differences here; but noteworthy are: critical ratios of 2.5 and 2.6 for C_H–P_m on the schizophrenia and hypomanic scales respectively, with C_H in each case the higher score; in the C_H–P_s comparison we find the respective critical ratios of 2.6 and 2.2 on the scales for depression and masculinity-femininity, with the P_s score higher on the depression scale, and C_H higher on the masculinity-femininity scale; while for C_H–P_{sv} on the masculinity-femininity scale the critical ratio is 2.1, C_H being the higher score.

Table 6 compares the P_m group with the P_s and P_{sv} groups. In every instance on the main scales the higher score is made by the more disturbed group. Significant differences are seen in the P_m–P_s comparison on all scales, the critical ratios being respectively 3.1, 4.1, 3.0, 3.6, 3.1, 3.8, 3.6 and 3.3 for hypochondriasis, depression, hysteria, psychopathic deviate, masculinity-femininity, paranoia, psychasthenia, schizophrenia, and hypomania.

In the comparison of P_m–P_{sv}, only on the schizophrenia scale is a significant difference apparent, the critical ratio being 3.0, with P_{sv} the higher score; but to be noted are the critical ratios of 2.5, 2.3, 2.7, 2.0 and 2.5 on the scales for depression, psychopathic deviate, paranoia, psychasthenia and hypomania respectively—all scores being higher for P_{sv}.

Table 7 compares the P_s–P_{sv} clinical groups, where no significant differences can be observed. However, it is interesting to note that the trend is for the P_s group to score higher on the scales for hypochondriasis, hysteria, masculinity-femininity, and psychasthenia; while the P_{sv} group is the higher on the scales for depression, psychopathic deviate, paranoia, schizophrenia and hypomania.

Summary

As indicated earlier this is not an experiment in the usual laboratory sense; yet, every effort was made to maintain standard conditions for all groups and to prevent the intrusion of uncontrollable variables or artifact. The value of the Multiphasic Inventory as a predictive instrument and clinical tool depends upon its agreement with the clinical diagnoses. This is, of course, open to many pit-falls: the lack of a common psychiatric language; the lack of clear-cut clinical criteria for identifying personality abnormalities; the failure of an individual under study to present uncomplicated symptoms or to present crystallized behavior patterns; the lack of time for making a prolonged study of the individual.

This explains to an extent the lack of greater definiteness in the diagnostic groups employed here. No attempt has been made to define the groups beyond the major classifications. Since a clinical evaluation had to be made with the minimum expenditure of time consistent with good procedure, and since verification of the clinical diagnoses was generally unpracticable, greater precision here would not mean necessarily substantiation in fact or in composite psychiatric or psychological opinion. Table 8 shows the recommendations made and the actual disposition on the 121 cases of this study. This is purely pragmatic. It must be borne in mind that final disposition of these cases rested with a board of medical or line officers, few of whom had any genuine training in psychology or psychiatry. Yet, this subjective system was in agreement with the clinical psychological-psychiatric findings minimally at 62.8%; 41.3% were discharged; 17.4% were placed on a limited duty status, and 4.1% were sent to a general hospital for further observation and disposition. The follow-up data are deficient in 29.8% of the cases studied. These men were returned to duty, but the percentages could not be ascertained as to which were for full duty and which for limited status. All of the psychiatrists making diagnoses were qualified neuropsychiatrists. Two of them were fellows in the American Psychiatric Association. Three of these showed little or no interest in the Inventory (or any psychometrics beyond intelligence examinations); the fourth neuropsychiatrist who had been on duty at the Consultation Service was definitely interested in the Inventory but purposely refrained from consulting it until after he had made a clinical evaluation.

The several profiles of Figures 1–5, and Tables 3–7 indicate the agreement of the test with the clinical findings and point out its potentialities as an aid in differential diagnosis. Objective comparison of the diagnostic groups with a normal group has demonstrated statistically significant differentiations (Table 3). Even between the diagnostic groups, one finds distinctive and valid differences (Tables 4–6), typifying, or peculiar, to a general class of disorder or diagnostic group. Thus, the depression scale appears as an indicator of the seriousness of the disintegration; being lower than both hypochondriasis and hysteria in the psychopathic group and mild neurotics, and higher in the severe neurotic and psychotic groups.

The lack of statistical significance between the P_s group and the P_{sy} group (Table 7) may be due to the small number of cases in the P_{sy} group and its large standard deviations. Qualitatively, however, the profiles are different (Figs. 4 and 5). It is to be noted that the peaks for hypochondriasis, hysteria and psychasthenia are more pronounced than the other scales for the P_s group; while, in the P_{sy} profile the peaks are for depression, paranoia and schizophrenia.

Table 8

Recommended Action and Final Disposition of the C_p, C_H, P_m, P_s, and P_{sy} Groups

		Discharge		Limited Duty		Full Duty		Admit. to Gen. Hosp.		Returned to Duty: Type Unknown			
		n	% within Group	n	% within Group	n	% within Group	n	% within Group	n	% within Group	n	%
C_p	N.P.												
	Recomm.	6	54.5	2	18.2	2	18.2	1	9.1	—	—	11	100
	Actual	3	27.3	1	9.1	0	0	1	9.1	6	54.5	11	100
C_H	N.P.												
	Recomm.	7	100	0	0	0	0	0	0	—	—	7	100
	Actual	6	85.7	0	0	0	0	0	0	1	14.3	7	100
P_m	N.P.												
	Recomm.	2	7.7	13	50	10	38.5	1	3.8	—	—	26	100
	Actual	2	7.7	5	19.3	1	3.8	0	0	18	69.2	26	100
P_s	N.P.												
	Recomm.	48	75.0	1	1.6	5	7.8	10	15.6	—	—	64	100
	Actual	32	50	14	21.9	6	9.4	3	4.7	9	14.0	64	100
P_{sy}	N.P.												
	Recomm.	10	76.9	1	7.7	0	0	2	15.4	—	—	13	100
	Actual	7	53.8	1	7.7	2	15.4	1	7.7	2	15.4	13	100
Totals	Recomm.	73	60.3	17	14.1	17	14.1	14	11.5	—	—	121	100
(of all cases)	Actual	50	41.3	21	17.4	9	7.4	5	4.1	36	29.8	121	100

The more probable explanation for the apparently close relationship between these groups (and also between C_p–P_{sv} and C_{II}–P_{sv}) is the grossness or overallness of the classifications used here together with the generally non-crystallized reaction types.

That the test is sensitive is further suggested in the N_o group. The profile for this group is generally slightly below the T-score mean of 50, but has a peak on the paranoia scale. In terms of its own gestalt, this would indicate a slight reduction in affect and a slight fixation of ideas. When one considers (cf. Table 1) that this group has been in the Service somewhat longer than the clinical groups and that they have met with more success (attained higher ranks) than the other groups, one is intrigued by the N_o profile. It is essentially stable and essentially normal. Yet, it does differ from the expected T-score average and from the normal profile as found by Leverenz (8). Could this be the healthy, experienced soldier's profile—phlegmatic and purposeful—or has there developed in this group a subtle personality alteration?

Provocative, too, of further investigation in the N_o group are the slightly elevated L and F scores. Has this group learned to put its best foot forward, but within affectively controlled limits, and at the same time has it become possessed with an idea, has it a singleness of purpose?

Conclusion

The purpose of this paper has been to present some findings on the use of the Minnesota Multiphasic Personality Inventory in a clinical situation. The cases were all male white enlisted men of the Air Force who had been studied at the Consultation Service of a personnel replacement pool. Of the cases seen, 98 were problems solely of administration and became the normal group. A total of 121 cases was seen because of maladjustment to the Army situation and these were subsequently diagnosed through standard practices by qualified neuropsychiatrists as either: constitutional psychopathic state, inadequate personality; constitutional psychopathic state, sexual psychopathy; psychoneurosis, mild; psychoneurosis, severe, or psychosis. Diagnoses were made independently of the Inventory. The data thus gathered indicate that the Multiphasic Inventory did in this empirical situation:

1. Distinguish graphically and with statistical significance between normal soldiers and those diagnosed as constitutional psychopaths; mild or severe neurosis; and psychosis.
2. Differentiate with significance between major clinical groups.
3. Present qualitative differentials, or hints for clinical query, in the more disintegrated or anomalous personality disorders.

The data are in agreement with Leverenz' (8) observation that although the clinical impression may not be corroborated always by the scores, the clinician is made aware of one or more personality abnormalities that require evaluation.

Received October 2, 1944.

References

1. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule: I. Construction of the schedule. *J. Psychol.*, 1940, 10, 249–254.
2. McKinley, J. C., and Hathaway, S. R. A multiphasic personality schedule: II. A differential study of hypochondriasis. *J. Psychol.*, 1940, 10, 255–268.
3. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule: III. The measurement of symptomatic depression. *J. Psychol.*, 1942, 14, 73–84.
4. McKinley, J. C., and Hathaway, S. R. A multiphasic personality schedule: IV. Psychasthenia. *J. appl. Psychol.*, 1942, 26, 614–624.
5. McKinley, J. C., and Hathaway, S. R. The Minnesota multiphasic personality inventory: V. Hysteria, hypomania and psychopathic deviate. *J. appl. Psychol.*, 1944, 28, 153–174.
6. Hathaway, S. R., and McKinley, J. C. *Manual for the Minnesota multiphasic personality inventory.* Minneapolis, University of Minnesota Press, 1943.
7. Hathaway, S. R. The personality inventory as an aid in the diagnosis of psychopathic inferiors. *J. consult. Psychol.*, 1939, 3, 112–117.
8. Leverenz, Major C. W. Minnesota multiphasic personality inventory: An evaluation of its usefulness in the psychiatric service of a station hospital. *War Med.*, 1943, 4, 618-620.
9. McKinley, J. C., and Hathaway, S. R. The identification and measurement of the psychoneurosis in medical practice: The Minnesota multiphasic personality inventory. *J. Am. med. Ass.*, 1943, 122, 161–167.
10. Schiele, B. C., Baker, A. B., and Hathaway, S. R. The Minnesota multiphasic personality inventory. *Journal-Lancet*, 1943, 63, 202–207.

Use of the Minnesota Multiphasic Personality Inventory in Vocational Advisement *

Lindsey R. Harmon and Daniel N. Wiener

Veterans Administration, Minneapolis, Minnesota

At each of the 53 regional offices of the Veterans Administration the Advisement and Guidance Subdivision in the Vocational Rehabilitation and Education Division determines the need for rehabilitation of disabled veterans and provides vocational counseling services leading to the selection of appropriate employment objectives. The work of rehabilitation, including advisement and training, is conducted in accordance with instructions developed by the Central Office of the Veterans Administration in Washington, D. C., to effectuate the provisions of Public Law No. 16, 78th Congress, as amended. The vocational advisers in the Subdivision use army or navy records, case histories, interviews, and vocational information for their contributions to an appropriate vocational choice. In addition, the results of valid, reliable, and well-standardized tests play a considerable role in the advisement, and are administered as deemed necessary in the advisement process. This advisement process has been described by Brophy and Long (1).

At the Minneapolis facility, the Advisement and Guidance Subdivision uses an extensive test battery which includes the Minnesota Multiphasic Personality Inventory as a measure of personality. The need for a personality test in vocational advisement is recognized as particularly acute since over half of the medical discharges from our armed forces have been for neuropsychiatric disorders. However, these disorders have been diagnosed only within the unique and tense framework of life in the armed forces and need to be related to the framework of civilian society. The Minnesota Multiphasic Personality Inventory was included in the personality measures because it appeared well standardized in terms of various clinical categories that could be related to vocational fields.

After a preliminary interview in which the provisions of the law and the regulations concerning vocational rehabilitation are explained to him, the veteran is assigned to an adviser who interviews him in order to complete a comprehensive survey of his background. On the basis of this interview, the adviser indicates on a worksheet the appropriate tests to be administered by the psychometrist. The Minnesota Multiphasic Personality Inventory is usually indicated. The veteran then begins his testing. When it is done, he returns to his adviser who discusses employment objectives with him on the basis of the records, the previous interview, and the test results.

---

* Special acknowledgments are due to the Veterans Administration which has furnished the data for this article, and to the officials of the Veterans Administration whose criticisms have been invaluable in its preparation. However, the authors assume sole responsibility for the opinions expressed.

The elements that comprise a counseling interview form a complicated pattern from which it is difficult to isolate for evaluation specific kinds of data. However, it is apparent in many cases that a test result, a physical or mental disability, an aspiration, or a family ambition, indicates so important a characteristic of the individual's behavior that it dominates the interview and the choice of the employment objective. Traits bearing this relation to the choice of an occupation are not infrequently revealed by the Minnesota Multiphasic Personality Inventory, even when no single scale is sufficiently elevated to indicate a clinical abnormality. Consideration of the several scales of this test and their use in the advisement of veterans in the Minneapolis office of the Veterans Administration during the first several months of the rehabilitation program will serve to illustrate the usefulness of the Multiphasic Inventory in this program.

The summary of interpretations and uses of the inventory in vocational counseling that follows is offered as the accumulated experience of the Rehabilitation and Education Division of the Minneapolis office of the Veterans Administration. It is the outgrowth of clinical experience rather than the statistical results of controlled experiments. It is hoped that experimental data may soon become available; pending such results, it is felt that these interpretations may serve as a useful guide in the employment of the test.

Use of the Scales

Elevation on the scales of Hypochondriasis, Depression, and Hysteria has occurred more frequently than on the other scales of this test. Such elevations often indicate severe limitations on the kinds of work a man is willing to undertake, most frequently eliminating dirty or heavy jobs. Also, it often indicates the desirability of a job at the lower end of the man's ability level, where a minimum of stress will be encountered; and the avoidance of such high-pressure occupations as selling, investigational, or promotional work. Similarly, when Paranoia and Schizophrenia scales are elevated, the relatively routine, well-regulated job is preferable to work requiring initiative, self-discipline and social contact, as these functions are particularly difficult for those with this withdrawn type of personality. By contrast, elevation on the Psychopathic Deviate and Hypomanic scales indicates the type of personality most likely to adjust in jobs of a relatively undisciplined nature, where individual initiative and aggressiveness are at a premium, and which afford a maximum of variety in work processes, locale, or associates.

Elevation on the scales indicative of paranoia, psychasthenia and schizophrenia has been encountered less frequently than on the other scales. When elevation on these three psychotic scales does occur, a less favorable prognosis of vocational success is indicated than would accompany elevation on the other six scales. A simplified program of training and employment appears advisable, with avoidance of occupations where the skillful maintenance of social relations is an important factor in success. Outstanding in such personalities are the less desirable, less utilizable forms of adjustment than may be associated with elevations on the other scales. Hypochondriacs, depressives, and hysterics show symptoms with which associates may sympathize; in the psychopathic deviates and hypomaniacs the deviations may be self-compensating or even rewarding in certain occupations; the effeminate man may be the best-adjusted one in a number of professions. But the paranoid, schizophrenic and psychasthenic forms of adjustment have no apparent advantages or compensations to balance the disapproval and penalties that are placed upon them. These adjustment-patterns appear to be obstacles to success in practically all occupations and require the poorest prognosis of vocational success.

The Masculinity-Femininity score is often the most difficult to interpret. It is sometimes depressed considerably below obvious manifestations of effeminate behavior and interests, and it is often high without any overt manifestations or expressed interests to confirm it or give it significance in terms of vocational advisement. There may be a dichotomy between feminine interests that can be acceptably and even successfully related to social relationships and occupational fields, and femininity or inversion of a socially unacceptable nature. In spite of these interpretative difficulties, this scale is frequently a useful guide to the counseling interview, indicating aesthetic propensities that can be related directly to a vocational field, or an effeminacy that contra-indicates vocational fields where tough-mindedness is desirable.

Test Battery Employed

In order to illustrate a few of the many ways in which the Multiphasic Inventory has proved useful in vocational advisement, a brief description of the test battery of which the Multiphasic Inventory is a part will be presented, and case histories will be examined. Each case is purposely selected to be typical of a group of cases; unique or spectacular illustrations have been avoided. Although the Multiphasic results are described more fully here than is the information provided by the other tests, the latter must not be and have not been overlooked in counseling.

The test battery that was administered to the cases to be described was as follows: Academic ability: Unit Scales of Aptitude, Vocabulary and Analogies, Otis Self-Administering Tests of Mental Ability, and the Ohio State University Psychological Test, Form 21, for college ability. Clerical aptitude: Minnesota Vocational Test for Clerical Workers. Mechanical aptitude: Revised Minnesota Paper Form Board Test, Minnesota Spatial Relations Test. Manual dexterity: Minnesota Rate of Manipulation and O'Connor Test of Finger and Tweezer Dexterity. Vocational preferences: Kuder Preference Record, Strong Vocational Interest Blank for Men (Rev.). Personality: Minnesota Multiphasic Personality Inventory. All norms are in terms of general late adolescent or adult populations, except where otherwise specified.

Scores, when given, are standard scores, except for the Kuder Preference Record which is in percentiles. The abbreviations used in reporting the Minnesota Multiphasic Personality Inventory scores are: ?, Cannot Say; L, Lie; F; Validity; Hs, Hypochondriasis; D, Depression; Hy, Hysteria; Pd, Psychopathic-Deviate; Mf, Masculinity, Femininity; Pa, Paranoia; Pt, Psychasthenia; Sc, Schizophrenia; Ma, Hypomania. For detailed information on the Minnesota Multiphasic Personality Inventory see references in the bibliography.

Illustrative Cases

L. S. This high school graduate is 22 and single. He had been in the Navy for 4 months, 3 months in training to be a radio operator, and 1 month in the hospital for serious delusions, depression, and anxiety. He was discharged for psychoneurosis rated 10% disabling. He applied for rehabilitation because his present work as a shipping clerk does not challenge him or offer a future, because he does not have the "background or amount of confidence" to fit him for other work, and because he cannot do work "calling for quick thinking or courage." He primarily wants to train for a career in music, secondarily to manage an auto parts or some other business. He states, "I think my will can overcome the difficulties" which might interfere with success.

On all tests but two, his academic, mechanical, clerical, and manual abilities and aptitudes are well above average. His highest points are the clerical tests, at the 99th and 97th percentiles. His preferences are for literary and musical work, with the Clerical-Computational pattern secondary. The Multiphasic Personality Inventory scores are: ? 50, L 50, F 80, Hs 65, D 70, Hy 67, Pd 86, Mf 98, Pa 67, Pt 92, Sc *99*, Ma 63.

This man was completely ineffectual in the interview situation. He looked down at the floor continuously, would respond only minimally and in a whisper, and was entirely amenable and suggestible to the adviser's words. He used abstract terminology, showed no aggressiveness, and in general lacked orientation to the real world. The Multiphasic Personality Inventory revealed enough extreme psychotic tendencies to warrant hospitalization at worst, and at best, an employment objective of low enough level to place him under minimal stress, yet respectable and clean enough to avoid irritating him with its physical demands.

There was no evidence of musical ability except for expressed interest in song writing. He showed some appreciation of the difficulties involved in making a living in the field of music, and the desirability of being able to earn a living in a more stable field. Bookkeeping appeared to meet his needs and abilities best and was entered as his objective. As soon as possible, this man should be referred for psychiatric aid. The strong schizophrenic trend indicated by the Multiphasic Personality Inventory and the interview situation make continuing success in any employment objective doubtful, although he is doing excellent work in his training at present.

J. M. This 22 year old veteran is a high school graduate. He came in to apply for rehabilitation because he was not qualified for any skilled job, and his present work, being unskilled, offered no future security. He had no expressed vocational preferences except for "desk work" which he "could do even if (he) didn't want to." His disability rendered his left arm useless for moderate or great muscular exertion.

His academic ability was above the average of the general population but below the average of college freshmen; his mechanical aptitude and manual dexterities were generally above average. His tested preferences were for persuasive, artistic, and literary work. The Minnesota Multiphasic Inventory scores were: ? 50, L 50, F 55, Hs 65, D 53, Hy 60, Pd 50, Mf *69*, Pa 50, Pt 66, Sc 67, Ma 68.

The first interview had revealed no evidence of the deviate personality tendencies which were later indicated by the Multiphasic. After the Multiphasic results were available, however, he was questioned more closely about his training, experience, and interests, in an attempt to discover the significance of the elevation in Mf. The Preference Record was also available, indicating literary and artistic interests, which were associated with his Mf score.

The second interview revealed not only some artistic interest, but actual commercial art school training and extensive spare time work at cartooning. He was reluctant to mention this background, he said, because he had been unable to finish paying the art school. The adviser's impression was that he also wanted to avoid appearing effeminate or artistic.

He later brought in his work and it was judged favorably by a commercial artist. To objectives suggested before he mentioned his art work, he had raised minor objections; later when an objective in art was suggested, he showed considerable relief. He was entered into training as a commercial artist where he is very happy and conscientious with his work. His chances for success appear excellent.

J. J. J. J. is a 29 year old man, with ninth grade education. He was a cook in the army and has a long work record that includes a variety of semi-skilled and unskilled jobs mainly on farms and in lumber camps. He feels he needs schooling to qualify for light work which he has been told is all he can handle now, and considers his present job as a Railroad Ashman too insecure. His rated disability is valvular heart disease, 80% disabling; he is in the class of heart disease ranked next to the completely bedridden on a five point scale.

Test results indicate a man of low academic ability, average clerical ability, and high mechanical aptitude and manual dexterities. His tested preferences are for mechanical and artistic work, with the clerical field secondary. The Multiphasic Inventory scores were: ? 50, L 70, F 50, Hs 53, D 70, Hy 75, Pd 47, Mf 51, Pa 41, Pt 39, Sc 40, Ma 43.

High point on the Multiphasic was Hysteria, with the Lie score and Depression also elevated. During the interview this veteran showed considerable concern about getting a "good deal," particularly looking for security. His only objection to his present work was that it was too insecure; actually it is heavy work that he should not be able to do if his disability were purely organic. This evidence together with the Multiphasic Inventory evidence prompted a discussion with the cardiologist in which he stated that the disability was probably not as bad as rated on the basis of the evident organic symptoms.

With this information, the advisement was subject to less limitation and it was possible to discuss with the veteran a fairly broad range of jobs in the mechanical field that involved up to a moderate amount of physical activity. The objective agreed upon was that of office machine serviceman involving a moderate amount of physical exertion which was not originally considered medically feasible.

There was considerable evidence of emotional disturbance in this man's frequent job changes, his indecision about what occupation he now wants to enter, the incompatibility of his present work and his disability, and his desire for education without specific reason. The Hy, D and L scores of the Multiphasic Inventory were valuable indicators of the form of his maladjustment.

J. G. This 20 year old unmarried veteran left school while in the tenth grade, joining the Navy as a fireman to begin what he planned to be a Navy career. He was discharged with pulmonary tuberculosis for which he receives maximum compensation. He began his advisement very certain that he wanted only to be a bookkeeper, and stated that he saw no use in the testing and advisement for him as he knew what he wanted and would fight until he got it.

Test results revealed above average academic ability, high mechanical aptitude and somewhat above average manual dexterities. On the clerical tests, he scored at the 9th percentile on Numbers and at the 31st percentile on Names. His interests were Computational-Clerical primarily, and Mechanical and Persuasive secondarily. The Multiphasic scores were: ? 50, L 50, F 53, Hs 53, D 51, Hy 58, Pd 47, Mf 53, Pa 35, Pt 41, Sc 47, Ma *63*.

Although the Ma score is not high in terms of absolute norms, it is high relative to the individual pattern, and is indicative of the strongest characteristic in this man's behavior. His behavior during his hospitalization and since his discharge confirms the Ma score. He practically forced his way out of the hospital, prematurely according to the doctors, and has insisted on immediate action throughout his dealings with the Veterans Administration, often to his disadvantage.

This veteran did not show sufficient clerical aptitude to render bookkeeping a desirable employment objective, and the personality inventory was higher on overproductivity of thought and action than one would associate with successful bookkeepers; nor was there evidence from the background or interview, of basis for his firm decision. The decisiveness of his decision was a function of his personality. The interview confirmed the Multiphasic Inventory, and it proved futile to attempt to find a rational basis for his preference, even though it was apparent that sales or mechanical work would be a more appropriate choice. Because of his insistence, he was permitted entrance into training as a bookkeeper. Prognosis for success in training was doubtful, and already reports on his progress indicate that he is "not putting forth his best effort," which he attributes to the effect of the weather on his lungs.

E. M. This single, 28 year old Negro veteran suffered a foot infection that required several months hospitalization and left him with limitations of motion of his foot and a need to avoid strenuous work. Shortly before discharge from the service he had been removed as platoon leader because of disobeying army regulations.

An intelligent, pleasant man, he made an excellent first impression which was strengthened by the results of aptitude tests showing good abilities for academic and mechanical work. He had previous college training in chemistry and wished to be trained as a pharmacist. His mediocre grades in college he attributed to the fact that he was working several hours a day as an electricians' helper in order to support himself at school. His vocational preferences were in the scientific, social service and clerical fields, thus fully supporting his expressed choice.

Because he completed his advisement early in the history of the program, before the administration of the Multiphasic Inventory had become routine, this test was not given to him. He was inducted into training with no reservations except those arising from the fact that, as a pharmacist, he might have to spend too much time standing on his injured foot, and the further limitations on placement possibilities because of his race.

His University grades were poor on his first examinations. It seemed reasonable to attribute his difficulty to the long interruption in his college training. When they continued poor, a check with his instructors revealed that his basic preparation in chemistry was poor and he did not seem to be making any progress in learning. He was referred back to the Advisement Subdivision with the notation that he apparently lacked the necessary mental equipment to handle college work. Concurrently, he came to the attention of the counseling service of the University, where—in addition to re-tests of scholastic aptitude and objective measures of achievement in academic subjects—the Multiphasic Inventory was given. The results were: ? 50, L 50, F 58, Hs 53, D 56, Hy 49, Pd *98*, Mf 63, Pa 70, Pt 70, Sc *74*, Ma 70.

A comparison of his report of his work and actual grades showed numerous discrepancies, and his reasons for poor previous achievement did not hold up under investigation. The clinical history, on further checking, yielded other data that completed the picture of a constitutional psychopathic deviate.

The re-valuation of his vocational potentialities in the light of the above facts made it apparent that he would not be able to achieve up to his mental ability because of his deficient self-discipline and direction. This line of reasoning resulted in a down-grading of his objective from professional work to the skilled labor of electrical-appliance repairing, thus utilizing maximally his previous experience, taxing his abilities minimally, and affording a fairly wide variety of activities on the job.

R. S. A veteran of the South Pacific campaigns, R. S. was discharged for mixed psychoneurosis, rated as 50% disabling, and in addition states he suffers from recurring malarial attacks. He is 24, and a high school graduate. His employment experience consisted of truck driving and helping in his father's laundry. He feels that he cannot return to any heavy work, is trained for nothing else, and would like to become a Forest Ranger or a Certified Public Accountant.

His tests show about average academic ability and clerical aptitude, and an erratic pattern on mechanical aptitude and manual dexterity tests, with scores more often above than below average. His tested preferences are for mechanical-scientific work, with computational work secondary. The Multiphasic Inventory scores are: ? 50, L 50, F 60, Hs 97, D 87, Hy 78, Pd 65, Mf 59, Pa 56, Pt 79, Sc 76, Ma 63.

Although he indicated a preference for forestry work, when discussion centered on specific physical demands, he stated that he could not do the manual work involved. His desire for it derived only from the possible college education involved. Neither did the idea of working alone in the forests appeal to him (note elevation on Pt). Mechanical work also, he was afraid, would be too trying for him physically. In spite of only average clerical aptitude, he would consider only bookkeeping as an alternative. This was entered as his objective on the basis of his expressed interests and his personality, since it involved a minimum of physical and emotional stress for him and was not strongly contraindicated.

The primary adjustment in this advisement had to be to the veteran's concern with his health, as the Multiphasic Inventory indicated; D and Hy are concomitants of Hs. Within the one month period since he was assigned to training, he had to postpone entering the program for several days and was absent for about four days since beginning. He is taking medicine for his throat and has complained of trouble with his eyes. The prognosis for his success is only fair, with the strong possibility that his extreme concern with his health will seriously interfere with his training and employment.

Summary

The Minnesota Multiphasic Personality Inventory, as part of the test battery used in vocational diagnosis of disabled veterans applying for rehabilitation, has proved to be an instrument of prime utility. It has served to delineate personality characteristics of crucial importance in the actual choice of a vocation, and has yielded valuable information to aid in prognosis of success in training. In some instances it has revealed personality characteristics that had not previously been recognized, and in others offered quantitative confirmation of the clinical impressions of a case history and interview.

It is most useful as a part of a well-rounded advisement procedure including tests of vocational aptitudes, mental ability, and vocational interests, and a thorough interview and case history. Improvement of specific scales, particularly the tests of internal validity, which are a unique and valuable aspect of this test, should further enhance its value. Additional clinical and experimental studies of the relation of test scores and patterns of scores to success in various vocations seem likely to extend the usefulness of this instrument in the vocational guidance clinic, but its practical value has already been sufficiently demonstrated to justify its employment on a basis of equality with tests of abilities and interests.

Received June 14, 1944.

References

1. Brophy, D. F., and Long, L. Veterans Administration vocational program. Processing procedures used by the College of the City of New York. *Psychol. Bull.*, 1944, 41, 795–802.
2. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule: I. Construction of the schedule. *J. Psychol.*, 1940, 10, 249–254.
3. McKinley, J. C., and Hathaway, S. R. A multiphasic personality schedule: II. A differential study of hypochondriasis. *J. Psychol.*, 1940, 10, 255–268.
4. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule: III. The measurement of symptomatic depression. *J. Psychol.*, 1942, 14, 73–84.
5. McKinley, J. C., and Hathaway, S. R. A multiphasic personality schedule: IV. Psychasthenia. *J. appl. Psychol.*, 1942, 26, 614–624.
6. McKinley, J. C., and Hathaway, S. R. The Minnesota multiphasic personality inventory: V. Hysteria, Hypomania and Psychopathic deviate. *J. appl. Psychol.*, 1944, 28, 153–174.
7. Hathaway, S. R., and McKinley, J. C. *Manual for the Minnesota multiphasic personality inventory.* Minneapolis: University of Minnesota Press, 1943.
8. Hathaway, S. R. The personality inventory as an aid in the diagnosis of psychopathic inferiors. *J. consult. Psychol.*, 1939, 3, 112–117.
9. Leverenz, Major C. W. Minnesota multiphasic personality inventory: an evaluation of its usefulness in the psychiatric service of a station hospital. *War Med.*, 1943, 4, 618–629.

First Impressions of Classmates

Wilhelmina E. Jacobson

Brooklyn College

That first impressions exert an important influence in human lives has long been recognized. Meeting new people, visiting strange places, facing unfamiliar situations, are part of the normal course of daily living, and in such circumstances first impressions are often the only guide to action. Where they are not the sole determinants of action, they unquestionably can contribute much towards specific behavior at the time or at future times. When first impressions are favorable, that is, when they elicit reactions anywhere from "love at first sight" to merely an unspecific sense of satisfaction, the natural result is acceptance or a desire for further acquaintance. Where the new situation takes the form of an interview for a job, for example, a favorable first impression on the employer would probably have the realistic result for the prospective employee of getting him the position. On the other hand, a negative first impression, leading to rejection and inclination to avoid further contact, would in the same situation mean failure.

While from a practical point of view some general knowledge of what tends to "make a good impression," or a poor one, is daily put to use in business and social relationships, yet very little is known specifically about first impressions. A study of their nature would therefore seem to be of considerable value, not only for its practical application, but for its value in helping to understand human behavior in general.

The college campus at the opening of the school year presented what appeared to be worth-while means of studying the nature of first impressions. A freshman class had just arrived. The setting was new, and first impressions during the early weeks of adjustment were discoverable in the choice of friends and of clubs, and the ease or difficulty of orientation to the college routine. Since social patterns are quickly set in new situations, the impressions of these early weeks were to have their influence, in all probability, on the entire school career of these students.

The writer had worked out, on a former occasion, a method for studying first impressions as a means of facilitating the orientation of a freshman *Clothing and Selection* class. Her own first impressions of the students as a group had been varied. To her they seemed hopeful, expectant, bewildered. As new students, their clothes, hair dress, and make-up were somewhat different from the campus mode. She believed that these students would be able to derive most benefit from the course if they could see themselves as others saw them and if they could learn to "take it."

It seemed that a technique used with success in one class might have value for the Personnel Department. As a result, four weeks after the opening of the term the present study, with 258 students in the freshman *College Problems* class as participants, was undertaken.

The Problem

The desire for approval, or to be accepted socially by one's associates, is a natural tendency of human beings. But what are the factors that cause one person to be accepted and another to be rejected? What must one be or do to be accepted by others? This study represents an attempt to explore this problem by the following two main lines of investigation:

I. *Determining the extent of favorableness of the impressions made* by freshman girls upon their classmates; II. Determining the nature of first impressions. In this phase of the study, two questions suggested themselves: *A*. What factors are involved? *B*. What influence do the observer's own characteristics have on her observations?

It was found that 22% of the 258 participants in the study had some degree of acquaintanceship with others in the class, ranging from "slight" to "being a friend." It seemed desirable to retain the data secured for this group and to see if they would reveal interesting comparisons between first impressions and impressions influenced by acquaintance.

The Group Studied. The subjects were women students in the College Problems class for freshmen at Ohio University. Of the 430 members of the class, 401 served as both observers (those who gave their responses) and as subjects (those about whom the responses were made) in the experiment. Twenty-two members helped to conduct the experiment; seven were absent. The subjects were typical college freshmen, women who came from various sections of the United States, but mostly from towns and farms of the state of Ohio.

Procedure. In a study of first impressions it is necessary that the subjects have no acquaintance with one another, or that their acquaintance be so slight as to be negligible. The 430 students were divided into 18 groups of approximately 24 students each. Twenty-two of the experimenter's college class in Clothing and Selection were put in one special group because they were to assist in the experiment. The groups were made up by choosing every seventeenth student from the alphabetically arranged roll. This was done in order to avoid acquaintanceships that the girls might have formed by sitting next to one another in classes during the five weeks since the opening of the term. Each group met from 11:00 to 11:50 o'clock in separate classrooms in two of the college buildings. The Dean of Women, who taught the weekly College Problems class as one unit, had told the girls in advance that they would meet that day in small groups. They were not informed as to the nature of this meeting, but were told to bring pencils as they would be asked to write. In this way the results obtained would be a cross section of student impressions on the campus on that particular day and at that hour.

The 22 students[1] who had been selected to help with the experiment were trained for their jobs in their regular two-hour laboratory period from 8:00 to 10:00 of the same day. During this period they were told for the first time the nature of the experiment to be conducted in the College Problems class that day, and were asked if they cared to assist. All volunteered. Each girl was assigned to a group; in five cases two students were assigned to a group.[2]

Three graduate students who were student deans also assisted in the experiment. Each student dean was assigned to one floor where the groups were assembled, to aid the student assistants if they needed help and to give encouragement. Each student assistant had the following materials: 1. A list of the names and special numbers of the 24 girls in her group; 2. Twenty-four packs of 3 × 5 index cards, 24 in each pack; 3. Typed directions (to be memorized); and 4. A watch with a second-hand, or a stop-watch.

The assistants were told to call the roll as soon as the bell rang, and to distribute the cards. They then asked the class: "Would you be interested in knowing what 23 girls would be frank enough to say about you? If so, each of you in turn will volunteer to come to the front of the room, state your name, and remain there while your classmates write down their impressions of you on these index cards."[3]

For each subject in turn the girls at their seats took a separate card, wrote the subject's number in the upper right-hand corner, and one of the following phrases in the lower left-hand corner: (1) "Don't know her," (2) "Know her slightly," (3) "Know her well," (4) "A friend." They were then given a minute and a half in which to list whatever remarks came to their minds about the subject as she stood before them at the front of the room. When every girl in the group had acted as a subject, the assistant asked the girls to write their own number in the lower right-hand corner of each card. The student assistants in the groups that finished before the end of the hour were told to instruct their girls to go back over their comments and to (1) place a plus (+) sign in front of the comment if they meant it to be favorable, (2) place a minus (−) sign if unfavorable, (3) place a zero (0) if it was in-between. Then each girl put a rubber band around her set of cards, after including a separate card on top which contained her own name and number.[4] All the cards were then collected.

---

[1] These students had been observers and subjects in the First Impressions experiment the first week of the term. They had studied the remarks made about them and were working on improvements. They had voted 100% in favor of using the same technique in subsequent Clothing and Selection classes.

[2] The assignment of these students to take charge of particular groups was based on the experimenter's knowledge of the students gained from results on the Bernreuter "Personality Inventory," from results on the Ohio State University Psychological Test, and from the students' work and cooperation in the Clothing and Selection class.

[3] Jennings says that "in order to secure valid data, it is evident that sociometric tests must hold reality value for the subjects to which they are administered." Helen Hull Jennings, *Leadership and isolation*, p. 28, 1943.

The Findings

As each girl stood before her classmates during the experiment, she elicited several very definite responses from her observers in the form of short comments or remarks, for example, "pretty blonde hair," "good posture," "very kind to others," etc. At the beginning of the experiment the aim was to study first impressions only, but when it was found that a 22% acquaintanceship existed among the students, as noted before, it seemed worth while to study impressions more intensively. The responses were therefore divided into four groups ("Don't know her"; "Know her slightly"; "Know her well"; "A friend") according to the extent of acquaintanceship. The term "first impressions" was reserved for one group only, the "Don't know her" group.

A total of 258 students (or 11 of the 17 groups)[5] completed the experiment and are included in the study findings. They served both as observers and as subjects. As observers they evaluated their responses as: + (favorable), 0 (in-between), and − (unfavorable). The acquaintanceship among these 258 students was 77% "Don't know her," 16% "Know her slightly," 3% "Know her well," and 3% "A friend."

The 258 individual observers gave a total of 19,352 responses, of which 12,149 were favorable, 1,903 were in-between, and 5,300 were unfavorable. This is an average (mean) of 47 favorable, 7 in-between, 21 unfavorable, and a total (mean) of 75 responses by each observer.

---

[4] The girls were told they need not sign their names on the top card if they were not interested in knowing what their classmates said about them. Of the 401 subjects, 306 signed these cards.

[5] The observers in six of the groups did not have time to make the evaluations, and are therefore not included in this report.

Stated in percentages, the responses were 63% favorable, 10% in-between, and 27% unfavorable.

The observers covered a wide range in number of favorable and unfavorable responses. Where one gave only 6 favorable responses, another gave as many as 105, or, stated in percentages, one gave 14% favorable responses, another 98%. The unfavorable responses ranged from 0 to 63 in number, or from 0% to 86%. One observer gave as few as 38 responses in all; another gave as many as 123.

These results show that the majority of the responses were favorable, the unfavorable responses being less than half the favorable. The observers tended to be definite in their responses, that is, either favorable or unfavorable rather than in-between.

To determine the nature of these students' impressions of their classmates, a detailed study was made of five of the eleven groups[6] by analyzing 9,076 responses made by the 116 girls who comprised the five groups.

The study was begun by first classifying the individual remarks made by the observers in three of the experimental groups being analyzed (these consisted of 3,335 individual responses given by 72 observers). With the aid of two seniors (this provided the student interpretation of student responses) the individual remarks were found to fall under 49 sub-topics. Many of the responses were found to have been repeated several times in more or less the same way. Therefore, to facilitate handling, the entire 3,335 remarks were grouped under 395 different responses under the 49 sub-topics.[7] The 49 sub-topics then seemed to lend themselves to further classification under five general categories: (1) physical characteristics; (2) intelligence; (3) clothing; (4) grooming; (5) psychological factors.

The next step was to obtain the consensus of expert opinion on the classifications. Four qualified members of the faculty made this evaluation: the Dean of Women, the head of the Department of Home Economics, the head of the Department of Psychology, and a member of the Department of Home Economics with a background in Psychology. They determined whether the 395 different responses belonged under the categories and sub-topics in which they had been placed, or elsewhere.

---

[6] A comparison of the selected five groups with the 11 groups with respect to mean Psychological Test Score, Bell Adjustment Inventory Scores, Degree of Acquaintance, and number and % of responses per observer, shows that they approximated the findings for the eleven groups as a whole, therefore making the study of additional groups unnecessary for the purpose of this experiment.

[7] Many of the sub-topics used by the experimenter and the two seniors were found to have the same meaning as those used by Gordon W. Allport. In these cases, therefore, Allport's terminology was accepted. Gordon W. Allport, *Personality*, p. 403, 1937.

Following is the final classification of categories and sub-topics determined on the basis of these judgments, with a typical response for each sub-topic:

Classification of Impressions, with Typical Responses

I. Physical Characteristics		
(A) Physique		
1. Hair	+	"Pretty blonde hair"
2. Eyes	+	"Large eyes"
3. Eyebrows	0	"Eyebrows could be thinned"
4. Mouth	−	"Not very attractive mouth"
5. Nose	+	"Cute nose"
6. Teeth	+	"White teeth"
7. Complexion	−	"Too pale"
8. Dimples	+	"Beautiful dimples"
9. Figure	+	"Good figure"
10. Height	−	"Too tall for a girl"
11. Weight	0	"Should lose some weight"
12. Posture	+	"Good posture"
13. Carriage	+	"Carries herself nicely"
14. General characteristics	+	"Very attractive face"
15. Masculinity-femininity	−	"Boyish in appearance"
(B) 16. Health	−	"Unhealthy looking"
(C) 17. Vitality	−	"Looks worn out"
(D) 18. Voice	+	"Nice voice"
II. Intelligence		
19. Abstract	+	"Looks brilliant"
20. Practical	+	"Looks efficient"
III. Clothing		
22. Harmony of parts of costume	−	"Should not wear heels with ankle sox"
23. Harmony of colors	−	"Too many different colors"
24. Suitability of clothes to person	+	"Wears becoming clothes"
25. Suitability to occasion	+	"Wearing suitable school clothes"
26. Manifestations of taste in dress	+	"Good taste in dress"
27. Fit of dress	−	"Skirt doesn't fit properly"
28. Remarks about clothing	+	"Pretty sweater"
29. Principles of art involved	−	"Bow is too large for its position"
IV. Grooming		
30. General remarks	+	"Dresses neatly"
31. Make-up	−	"Needs more make-up"
32. Hair dress	−	"Hair could be fixed differently"
33. Shoes	−	"Shoes could be cleaned"
34. Nails	−	"Too dark a shade of nail-polish"
35. Cleanliness	+	"She looks clean"
36. Laundering	−	"Belt is a little soiled"

V. Psychological Characteristics

21. Manifestations of emotions	–	"Afraid to smile in public"
37. Ascendance—sub-missive	+	"Dominant at times"
38. Expansion—reclusion	0	"Quiet"
39. Persistence—vacillation	–	"A bit flighty"
40. Extroversion—introversion	0	"Acts as if she would be nice after you know her but seems to have a wall you would have to break through first"
41. Self-objectification—self-deception	+	"Imagine she would be sincere"
42. Self-assurance—self-distrust	+	"Seems rather sure of herself"
43. Gregariousness—solitariness	+	"Very sociable or friendly"
44. Altruism (socialization)—self-seeking (unsocialized)	+	"Very kind to others"
45. Social intelligence (tact)—low social intelligence (tactless)	–	"Not always tactful"
46. Directed toward values	+	"Seems serious about college"
47. Radicalism—conservatism	+	"Very conservative type"
48. Observers' likes and dislikes of subject	–	"Don't think I'd like her"
49. General responses	+	"A great deal of charm"

A tabulation was then made of the 9,076 responses of the 116 observers. These were classified according to categories, sub-topics, observers, favorableness, and extent of acquaintanceship. It was thus possible to study impressions from many points of view.

What Factors are Involved in the Observers' Responses?

The groups or sub-topics under which the responses were classified show wide variation in frequency of mention, and also in extent of favorableness. Six of the 49 sub-topics constitute over 50% of the responses: Grooming—general remarks; grooming—hair dress; posture; emotions; self-assurance—self-distrust; altruism—self-seeking. Eighteen sub-topics include 93% of the responses (as indicated in Table 1). These 18 sub-topics also include 1% or over of the total of 9,076 responses. The remaining 31 sub-topics consist of less than 1% each.

The sub-topic of greatest frequency concerned general remarks on grooming, pertaining mainly to neatness. These general remarks and remarks on hair dress (second in frequency) cover 23.7% of all responses; when posture (which is third) is added, they make up 33.9%, or one-third of all the responses. Add the next three sub-topics—emotion, self-assurance—self-distrust, and altruism—self-seeking, and over 50% of all the responses are covered.

The five sub-topics which rank highest in favorable responses are: 1, hair; 2, eyes; 3, altruism—self-seeking; 4, general physical remarks; and 5, taste in dress. For each one of the sub-topics the favorable responses were over 85%.

Table 1

Number and Percentage Distribution of the 18 Highest Ranking Sub-Topics

Rank	Sub-topics No.	Sub-topics Item	Responses No.	Responses % of (9,076)	Evaluation of Responses in Per Cent +	0	−
1	30	Grooming—general remarks	1,224	13.5	82	4	14
2.5	32	Grooming—hair dress	929	10.2	47	10	43
2.5	12	Posture	920	10.2	51	14	35
4	21	Emotions	797	8.8	64	9	27
5	42	Self-assurance—self-distrust	582	6.4	13	27	60
6	44	Altruism—self-seeking	526	5.8	89	3	8
7	43	Gregariousness—solitariness	509	5.6	81	6	13
8	26	Taste in dress	432	4.8	87	6	7
9	14	General characteristics—physical	413	4.6	88	7	5
10	31	Grooming—make-up	378	4.2	17	12	71
11	25	Clothing—suitability to occasion	240	2.6	72	3	25
12	23	Clothing—harmony of color	235	2.6	57	4	39
13	1	Physical characteristics—hair	230	2.5	97	1	2
14	24	Suitability of clothes to person	194	2.1	41	5	54
15	7	Physical—complexion	182	2.0	74	10	16
16	2	Eyes	145	1.6	95	0	5
17	22	Clothing—harmony of parts	139	1.5	31	5	64
18	28	*Remarks about clothing*	115	1.3	83	8	9
		Total	8,190	90.3			

When an analysis is made of the unfavorable remarks, it is obvious that four sub-topics—make-up, harmony of parts of costume, self-assurance—self-distrust, suitability of clothes to person—rank highest in that order, and all are over 50% of the responses in each sub-topic.

The procedure in this experiment was such as to produce (reasonable expectancy) self-consciousness in the individual subjects, yet 27% of the self-assurance—self-distrust responses were neither favorable nor unfavorable, and these ranked highest of the in-between responses.

Relationship between Categories and Acquaintanceship Groups

A study of Table 2 makes possible the following generalizations regarding relationships between categories and acquaintanceship groups.

1. Psychological responses constitute the greatest proportion of the total responses, 30% of the total number. They are also most frequent in all acquaintance groups, except the "Don't know her" group, where responses concerning grooming are slightly in the lead. This tends to indicate that grooming is most obvious to the stranger, but as the girls

Table 2

Percentage Distribution of Responses (116 Observers)

Category	Don't Know Her	Know Her Slightly	Know Her Well	A Friend	All Groups
Physical	24.0	24.2	27.0	24.8	24.2
Intelligence	1.1	.8	.6	.9	1.1
Clothing	15.8	14.8	12.2	13.3	15.4
Grooming	30.3	27.2	22.7	22.6	29.3
Psychological	28.8	33.0	37.7	38.5	30.2
Total	100.0	100.0	100.1	100.1	100.2
	N = 6,899	1,558	393	226	9,076

become better acquainted, responses of a psychological nature increase, this increase, in this study, being 9.7%. They seem to indicate that strangers tend to comment about grooming, while acquaintances tend to comment concerning traits, attitudes, and temperament.

2. Grooming responses take second place and show a decrease as acquaintanceship increases, this decrease being 7.7%, namely, from 30.3 to 22.6%.

3. The responses of a physical nature take third place in "All Groups" and in the "Don't know her" and "Know her slightly" groups, and second place in the last two groups. The better acquainted the girls were, the more they tended to comment about physical characteristics and less about clothing and grooming. Responses of a physical nature also remain almost constant for each acquaintance group, namely, about 24%.

4. Clothing takes fourth place in all acquaintance groups. The clothes the students wore elicited comment only half as frequently as grooming or psychological characteristics. Clothing responses remain nearly the same for each group, ranging from 12 to 15%.

Relationships between categories and acquaintanceship groups, as influenced by favorableness, are indicated in Table 3. Certain inferences can be drawn from the percentages given.

1. A large majority of the 9,076 responses indicate favorable impressions—62% favorable to 28% unfavorable. The proportion is five favorable responses to three unfavorable and in-between combined. This majority of favorable responses is true for the total responses in all five categories. This finding corroborates Jennings' statement that "positive expression for participating with others was practically twice as great as the expression of rejection."[8]

---

[8] Helen Hall Jennings, *Leadership and isolation*, p. 59, 1943.

Table 3

Per Cent of Favorable and Unfavorable Responses Classified According to Acquaintanceship Group and Category (116 Observers)

Category		Don't Know Her	Know Her Slightly	Know Her Well	A Friend	Total
Physical	+	67.5	70.0	84.0	91.1	69.3
	0	9.9	8.0	3.8	1.8	9.0
	–	22.7	22.0	12.3	7.1	21.7
	Total	100.1	100.0	100.1	100.0	100.0
	N =	1,653	377	106	56	2,192
Intelligence	+	74.4	69.2	100.0	100.0	74.7
	0	14.1	23.1	00.0	00.0	14.7
	–	11.5	7.7	00.0	00.0	10.5
	Total	100.0	100.0	100.0	100.0	99.9
	N =	78	13	2	2	95
Clothing	+	62.3	69.7	77.1	86.7	64.6
	0	5.5	5.2	00.0	3.3	5.2
	–	32.2	25.1	22.9	10.0	30.2
	Total	100.0	100.0	100.0	100.0	100.0
	N =	1,088	231	48	30	1,397
Grooming	+	56.5	64.5	74.2	76.5	58.7
	0	7.6	7.1	1.1	2.0	7.2
	–	35.9	28.4	24.7	21.6	34.0
	Total	100.0	100.0	100.0	100.0	99.9
Psychological	+	56.0	66.3	68.2	87.4	59.7
	0	13.3	11.7	10.1	5.8	12.6
	–	30.7	22.0	21.6	6.9	27.8
	Total	100.0	100.0	99.9	100.1	100.1
	N =	1,987	514	148	87	2,736
Total	+	60.1	67.3	75.1	85.8	62.6
	0	9.5	8.7	5.1	3.5	9.0
	–	30.4	24.1	19.9	10.6	28.3
	Total	100.0	100.1	100.1	99.9	99.9
	N =	6,899	1,558	393	226	9,076

2. Responses of a physical nature show a greater per cent of favorable responses (omitting intelligence) than clothing, psychological or grooming, as shown in the "Totals" column of Table 3, the percentage being 69.3.

The fact that comments regarding physical features were usually favorable became quite noticeable while reading the 19,352 responses. It is interesting to note that unattractive features of the face were rarely commented on. Most of the 21.7% of unfavorable comments in the physical group concerned posture (42.4%), a factor which usually can be corrected.

3. Clothing and grooming show the greatest per cent of unfavorable responses (30 and 34 respectively). These features can be changed; therefore the girls may have felt they could be critical.

4. The in-between responses show the greatest per cent of responses in the psychological category and the next greatest in the physical. Again, the general tendency of the observer is to refrain from responding, or to give an indifferent response, if the feature is one that cannot be easily changed.

5. An increase in the extent of acquaintanceship is accompanied by an increase in favorable impressions, as shown in the per cent of total responses, where the percentage increase is from 60.1 for "Don't know her" to 67.3 for "Know her slightly," 75.1 for "Know her well," and 85.8 for "A friend." This increase is true for each of the individual groups,—physical, clothing, grooming, and psychological.

6. The per cent of increase in favorable comments with acquaintanceship (that is, from "Don't know her" to "A friend") was sharpest in the psychological category,—31.4%. As the students became better acquainted they showed a greater tendency to remark favorably about psychological factors. Grooming shows the least gain, 20%.

What Influence do the Observer's Own Characteristics Have on Her Observations?

As noted before, the personality adjustment of each student was obtained from results on the Bell "The Adjustment Inventory" and her intelligence score on the Ohio State University Psychological Test. It seemed that certain inferences could be drawn from the observer's characteristics, as revealed by these tests, and her observations.

1. Correlations between the observers' scores on the separate sections of the Bell-Adjustment Inventory and the favorable responses made by the observers show coefficients less than .089, except in the case of Bell (Home) and favorable impressions where $r = .28$, a value significant at the .01 level. This seems to indicate a slight tendency for students with satisfactory home life to give favorable responses about their fellow-students.

2. The observer's personality adjustment (Bell-Total) and the kind of responses the observer makes—physical, clothing, grooming, or psychological—seem to be independent of each other. Each of the four correlations has a numerical value less than .13, and these are not significant at the .01 level.

What influence does the observer's intelligence have on the kind of response she makes?

1. The observer's intelligence (Ohio State University Psychological Test) and the favorable responses she makes seem to be independent of each other ($r = -.10$). The more intelligent observer, on the whole, is about as favorable in her responses as the less intelligent observer.[9]

2. The observer's intelligence and the kind of response she makes—physical, clothing, grooming, or psychological—seem to be independent of each other, except in the case of clothing where $r = .34$, being significant at the .01 level. This seems to indicate that there is a tendency for students who are above average in intelligence to make more than an average number of comments about clothing. College women on the freshman level are clothes-conscious. It is possible that the more intelligent girl may be more observant, and also that she may know more about the latest mode, or what is correct in costume, than the less intelligent girl.

3. The more intelligent observer tends to give a greater variety of responses than the less intelligent: $r = .32$, this being significant at the .01 level. This may indicate that the less intelligent student tends to make the same type of comment in giving her impressions, limiting her statements to a few sub-topics.

Certain other factors seemed to influence the observer's impressions.

1. Observers differ in their taste in dress. One remarked about a subject: "She dresses nicely" (+). Another stated of the same subject: "Anklets don't match dress" (−).

2. Observers differ in their standards of grooming. One remarked: "Well groomed" (+), while another's comment about the same subject was: "Could be more neatly dressed" (−).

3. Observers differ in their interpretation of traits and attitudes. One observer remarked: "Looks friendly" (+), while another said of the same subject: "Looks like she might be hard to get along with" (−).

4. Observers differ in their conception of "Good looks." One observer remarked about a subject, "Not attractive" (−), while another remarked: "She has simple beauty" (+), while a third described her as "Fairly attractive" (0).

Summary

This study of impressions of one's classmates shows that the freshman girls included in the study were favorable in their evaluations of their classmates rather than unfavorable or neutral. There were three favorable responses for every two unfavorable and in-between combined. This would indicate that the "frankness of modern youth" is of a kindly nature rather than the opposite.

---

[9] Buck found a low correlation between thinking and behavior. Buck, N. M., and Ojemann, R. H., Relation between ability in scientific thinking and behavior in situations involving choice. *J. exper. Educ.*, 1942, 11, 217.

The 9,076 responses made by the 116 observers were found to fit in the five selected categories: (1) physical characteristics; (2) intelligence; (3) clothing; (4) grooming; (5) psychological characteristics, such as traits, attitudes, and temperament. They also could be broken down into the 49 concrete factors of varying importance in the judgment of these freshman girls. However, over 50% of the impressions were covered by 6 sub-topics and 90% by 18 sub-topics (see Table 1).

Of the five categories, responses of a psychological nature were most frequent, with grooming second, physical characteristics third, clothing fourth, and intelligence fifth. Students who were not acquainted tended to remark on grooming, but as acquaintance increased they tended to comment increasingly on psychological characteristics. Degree of acquaintanceship has little effect on the per cent of responses of a physical or clothing nature.

Characteristics of a physical nature, especially the face and features, were commented on favorably more frequently than other factors. On the whole, the students were very considerate when nothing could be done to correct a defect. But they were equally critical when grooming or clothing was concerned.

There seemed to be no significant correlation between a student's own personality adjustment and her inclination to give favorable or unfavorable responses, or to comment about physical, grooming, clothing, or psychological factors. Where the girls had favorable home conditions, however, they tended to respond favorably.

Nor was there significant correlation between a student's intelligence and her favorable or unfavorable replies or her responses regarding physical, grooming, clothing, or psychological characteristics. They seem to be independent of each other, except in the case of clothing, where there appears to be a slight tendency for the more intelligent girl to comment about clothes.

The responses showed that students vary greatly in their taste in dress, their standards of grooming, in their interpretation of traits and attitudes, and in their criteria for "good looks."

The responses of a physical and clothing nature are about constant for the four acquaintance groups. But responses regarding grooming decrease as acquaintanceship increases, while psychological responses increase with acquaintanceship.

Within the limits and limitations of this study, the results may suggest the pattern of responses necessary to get along with other people. Although limited in scope, its findings may point the way to areas which need development.

Received February 9, 1944.

References

1. Barker, Roger G. The social interrelations of strangers and acquaintances. *Sociometry*, 1942, 5, 169-179.
2. Bonney, M. E. Personality traits of socially successful and socially unsuccessful children. *J. educ. Psychol.*, 1943, 34, 449–472.
3. Brogden, H. E., and Thomas, W. F. Primary traits in personality items purporting to measure sociability. *J. Psychol.*, 1943, 16, 85–97.
4. Buck, N. M., and Ojemann, R. H. Relation between ability in scientific thinking and behavior in situations involving choice. *J. exper. Educ.*, 1942, 11, 215–219.
5. Drake, M. J., and Others. Relationship of self-rating and classmate rating on personality traits. *J. exper. Educ.*, 1939, 7, 210–213.
6. Drought, N. E. Analysis of eight measures of personality and adjustment in relation to relative scholastic achievement. *J. appl. Psychol.*, 1938, 22, 597–606.
7. Durer, M. A., and Richards, G. V. *Appraisal of interpersonal relationships. Educ. Res. Bull.*, 1942, 21, 172–179.
8. Jennings, Helen Hall. *Leadership and isolation; a study of personality in interpersonal relations.* New York: Longmans, Green & Co., 1943.
9. Lunger, R., and Page, J. D. Worries of college freshmen. *Ped. Sem.*, 1939, 54, 457–460.
10. Richards, T. W., and Ellington, W. Objectivity in the evaluation of personality. *J. exper. Educ.*, 1942, 10, 228–237.

The Value of Aptitude Tests for Factory Workers in the Aircraft Engine and Propeller Industries *

John T. Shuman

Williamsport Technical Institute, Williamsport, Pennsylvania

The general problem of this investigation might be stated as follows: to develop a testing program for a company manufacturing radial and opposed types of aircraft engines, and to investigate the feasibility of using these same tests for upgrading the supervisory forces of three dissimilar companies of the same parent corporation.

Both factory and supervisory jobs were investigated at Lycoming Division—The Aviation Corporation. This industrial plant manufactures aircraft engines involving all the jobs necessary to the precision machining of engine parts. Only the supervisory jobs were investigated at a foundry, Spencer Heater Division—The Aviation Corporation, and at a propeller plant, The American Propeller Corporation.

More specifically the investigation was designed to determine whether tests at the Lycoming engine plant would:

a. Assist in the selection of new employees without previous experience in this type of industrial activity.

b. Assist in the selection of new employees for work in fields as machine operation, inspection, engine testing, the more skilled production jobs.

c. Assist in the selection of new employees to be trained as skilled workmen, as tool room workers, tool designers.

d. Assist in the problem of upgrading and promoting employees to supervisory positions.

e. Assist in the better adjustment of employees.

The investigation was further designed to determine whether these same tests utilized in the Lycoming plant could be used to assist in the problem of selection, and promotion of employees to supervisory positions in all three of the plants studied.

---

* An Abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Education at The Pennsylvania State College, February, 1944. The author wishes to acknowledge the helpful suggestions received from Dr. B. V. Moore, Dr. C. C. Peters, and Dr. F. H. Koos in connection with various phases of the investigation. A second article dealing with supervisors will follow in the next issue of *J. appl. Psychol.*

Criteria for Measuring Job Success

A job rating made by the worker's superior was utilized to measure success on the job.

Each workman was rated good, average or poor on his job. If he had been transferred from one job to another, the reason for the transfer was ascertained.

The employee's immediate supervisor was also asked what good and poor points the ratee exhibited. The reason for transfer and the comments were used as checks on the rating; if they did not parallel the rating, the supervisor was checked to determine the reason for the difference.

The criteria used were the best possible for this particular situation; and the work of securing the criteria was carefully executed. Ratings were secured by several trained workers who contacted personally supervisors familiar with the ratees. No rating was made unless one or more supervisors thoroughly familiar with the work of the ratee could be found. Because of "war conditions" inside the plants this was sometimes difficult and critical.

The reliability of the criteria was found in two situations. These checks were made by having a trained departmental instructor rate the operators, and then securing a rating through the production supervisor as explained above. The reliability coefficients secured were: Production Engine Testers, .91 (N-42); Inspectors, .705 (N-36).

Procedure

Test results were secured in two ways. Some groups of experienced operators were tested. Each of these operators had six months or more experience on a job, and it was assumed that he had proved satisfactory on that job or that he would not have been allowed to remain on it. Ratings were secured on these groups in the usual way after testing.

New applicants for jobs also were tested. Those applicants who were successful in securing jobs were followed up after six months or more in the plant, and ratings secured on them.

The following tests were used in the investigation: Otis Quick-Scoring Test of Mental Ability, Beta, Form A; Revised Minnesota Paper Form Board, AA; Bennett Test of Mechanical Comprehension, Form AA; O'Rourke Test of Mechanical Ability, Jr. Gr.; and Minnesota Vocational Test for Clerical Workers, Number Comparison.

The use of the O'Rourke test was abandoned early in the investigation because of its length, and because it proved to be almost useless with female applicants, many of whom simply didn't try to take it. Further, as more and more persons were hired from other than industrial occupations, the basis for this test seemed not to be suitable.

The value or effectiveness of the tests is assessed in terms of simple percentage figures indicating the extent to which their use increased the effectiveness of the current personnel procedures. That is, by how much will the tests increase the effectiveness of the present hiring job as shown by the percentage of good operators hired.

Although it is believed that the correlation does not give a true picture of the results, they are also presented in the form of biserial r's for those technicians who prefer to think in these terms.

Results are presented in a number of histograms and tables. Summaries of the tables are given here.

Table 1 presents a summary of the recommended minimum critical scores and the per cent improvement effected in the selection of excellent workers.

Table 1

Improvement Effected in Selection of Excellent Workers by Selected Minimum Critical Scores, Lycoming Division—The Aviation Corporation

	Otis Q.S. Test of Mental Ability, Beta A		Minnesota Paper Form Board, Rev.		Bennett Test of Mech. Comprehension, AA	
Job Group	Minimum Critical Score	% Improvement Selection, Excellent Workers	Minimum Critical Score	% Improvement Selection, Excellent Workers	Minimum Critical Score	% Improvement Selection, Excellent Workers
Toolmaker learners (N-64)	51	9	33	7	36	5
Inspectors (M & F) (N-49)	51	27	37	29	M-34 F-10	12 28
Engine *testers* (N-45)	46	19	21	—	33	10
Job setters (N-25)	34	17	30	30	36	47
Foremen (N-99)	35	9	24	8	30	10
Machine operators (M & F) (N-81)	31	10	21	8	M-27 F-18	22 12

Table 2

Average Improvement Effected in Selection of Excellent Workers, All Jobs, Lycoming Division[*]

Test	Mean Improvement in Selection Excellent Workers at Minimum Critical Score	Mean Improvement in Selection Excellent Workers at Q_1
Bennett Test of Mechanical Comprehension, AA	18.2%	19.0%
Otis Test of Mental Ability, Beta Test, Form A	15.1%	16.2%
Revised Minnesota Paper Form Board, AA	13.6%	14.0%

[*] The means given in the Table are not weighted by the N's but are averages of percentages by job categories. This was done because the mean of the different job categories and not of the number tested was desired.

Table 2 was developed by the simple expedient of averaging the results secured by each test for all job categories in the Lycoming Plant (see Table 1). Table 2 gives the mean per cent of improvement in selection effected at the various Minimum Critical Scores and at Q_1. These means indicate that the Bennett Test was on the whole the most effective of the three, with the Otis Test next and the Revised Minnesota Paper Form Board the least efficient.

Table 3

Correlations Between Job Ratings and Test Scores, Lycoming Division—The Aviation Corporation

	Biserial r		
Job Group	Otis Q.S. Test of Mental Ability, Beta A	Minnesota Paper Form Board, Revised	Bennett Test of Mechanical Comprehension, AA[*]
Inspectors	.52 ± .09	.50 ± .09	.665 ± .13
Engine testers	.57 ± .13	.16 ± .17	.17 ± .17
Machine operators	.48 ± .08	.38 ± .08	.44 ± .08
Foremen	.30 ± .07	.47 ± .07	.405 ± .07
Job setters	.46 ± .14	.59 ± .13	.73 ± .10
Tool room learners	.485 ± .09	.42 ± .09	.46 ± .09
Mean bis$_r$[**]	.49 ± .035	.44 ± .04	.52 ± .03

[*] All correlations in this column are for male only. The inspectors and machine operators are the only two categories affected. All other job groups have no female workers.

[**] The method used for averaging was that recommended by Garrett, *Statistics in psychology and education*, Longmans, p. 284. Each r was squared, the squares averaged, and the square root extracted of the average thus obtained.

Averaging the r's in Table 3 seems a justifiable procedure because the r's did not differ greatly in size and one of the objectives of this investigation was to determine to what extent certain tests exercised discrimination in a relatively large number of job classifications.

Conclusions

1. The three tests most extensively used selected excellent workers in the following order reading from the most to the least effective; that is, the Bennett Test was the most selective, the Otis next, and the Minnesota Paper Form Board the least efficient of the three.
2. Job-success on the purely manual-type job in the Lycoming plant shows no relation with the test scores. Discussion of these jobs has been omitted from this report.
3. Job-success on the jobs requiring skill, such as machining precision parts, testing aircraft engines, and inspection was found to relate positively and significantly with the test scores, in varying percentages or degrees.
4. The large range of aptitudes found in the industrial plant studied indicates that tests might best be used for assignment of employees to a job category or level rather than to specific jobs. In other words the function of a testing program should be primarily the better adjustment of workers.
5. The Minnesota Number Checking Test used in studying Production Engine Testers seems not to have a wide enough application on other types of work to warrant its use with other groups.
6. Under tight labor conditions tests in general must be usable with both men and women. A test such as the Bennett Test of Mechanical Comprehension, if usable with both men and women with little variation in range, would be much more valuable in many cases.

Received March 15, 1944.

Studies in International Morse Code

4. A Note on Second-Level Training in Code Reception

Fred S. Keller

Columbia University

In the first paper of this series,[1] there was described a new procedure for teaching beginners to recognize the thirty-six principal signals of International Morse Code. Essentially, this procedure involves the presentation of an auditory signal to the student and a subsequent identification of the signal by the instructor, with a short pause between signal and identification, to permit the student to write down the appropriate character (letter or digit). The signals are transmitted and identified individually; one hundred signals, arranged in random order, constitute a practice "run," after which the student is given a brief rest-period in which to sum his errors and enter the total in a "box-score" record.

Although no sharp line may be drawn between this early training in the accurate recognition of individual signals and the later practice in slow-speed reception of a series of signals, it is convenient to distinguish between the stage in which the fundamental stimulus-response relationships are established and that at which *speed* of response becomes all-important. The former may be termed "first-level," and the latter, "second-level," training.

The present report is concerned with two variations of a method of second-level training which have proved to be effective in bringing students of college age from a code speed of five words[2] per minute (5 w.p.m.) to speeds as high as 20 w.p.m. in less than fifty-five hours of instruction. In either form, this method differs from those commonly used in training centers, in that (1) it provides the student with knowledge of his progress at practically every moment throughout the entire period of instruction; (2) it encourages close attention to *all* signals transmitted within any practice session; and (3) it provides regular rest pauses at increasingly frequent intervals as the student advances in speed.

---

[1] Keller, F. S. Studies in International Morse Code. 1. A new method of teaching code reception. *J. appl. Psychol.*, 1943, 27, 407–415.

[2] A "word" is here treated as any five-character group of letters and digits randomly arranged—e.g., AR24P, B8GRX, etc.

Variation 1. The first variation of this method requires as many operators or automatic sending-machines as there are different speeds at which students within a group are qualified to work. In a class of twenty or thirty college students, with available code-speeds of 5, 7, 10, 12, 15, and 20 w.p.m., more than three instructors are seldom needed in order to meet the problem of individual difference in progress; but a larger or less homogeneous group of students might require the services of four, or possibly five, operators transmitting at as many different speeds.

Following the satisfactory completion of first-level training, the student is introduced to a 5-w.p.m. rate of transmission. Signals continue to be sent in runs of one hundred each, with all characters represented with equal frequency, in haphazard arrangement. The signals are not, however, individually identified by the instructor as in first-level training. Instead, at the end of each run, he "calls back" the phonetic equivalents (Able, Baker, Charlie, Dog, etc.) of the signals just transmitted, and the student checks, or corrects, his erroneous or omitted characters. The same practice sheets are used, and error totals are recorded in the same box-score form, as in first-level training (1, see Figs. 1 and 2).

At a 5-w.p.m. rate of transmission, a run of signals requires four minutes, and four or five minutes may be consumed in checking errors, as the characters are called back, and in recording scores. It is usually possible to transmit five runs within a fifty-minute practice session. At higher rates of speed, of course, a larger number of runs may be given, although the between-run interval remains fairly constant.

Students are advanced from one speed to another as soon as they have reached some arbitrarily chosen degree of proficiency. The results reported in this paper are based upon a mastery criterion of three successive runs in no one of which the student made more than five errors or omissions. This was also the criterion of first-level proficiency, as well as that in the procedure next to be described.

Variation 2. This variation of the calling-back method may be employed when a single instructor, working with a single sound source, is required to give instruction to a group of students. It is usually introduced while first-level training is still going on, and it differs from Variation 1 only in that it calls for the transmission of signals at different speeds to the same students during the same class hour. In order to accomplish this, the instructor may proceed, in stepwise fashion, from the slowest to the highest speeds at which his students are able to work, giving most practice at the speed suitable to the greatest number of students. Thus, when a few students have qualified at the first level of training, a single run of 5 w.p.m. may be introduced at the end of the class hour; when more students reach this level, more 5-w.p.m. runs may be added; when the 5-w.p.m. speed is mastered by some students, a 7-w.p.m. run may be introduced; and so on throughout the training period. As soon as all of the students, or all but one or two, have qualified at the lowest speed in use, then, of course, that speed may be omitted entirely.

Table 1

Progress of Two Groups of College Students in Receiving Morse Code under Two Variations of Second-Level Instruction

Speed (w.p.m.)	Group I (Variation 1) Hours* (50-minute)	Group I (Variation 1) Number of Students	Group II (Variation 2) Hours* (50-minute)	Group II (Variation 2) Number of Students
5	17.8	40	18	56
7	23.1	38	25.4	56
10	35.0	33	37.5	55
12	44.7	22	51.5	24

* The values in the "Hours" columns are cumulative. The decreasing number of students at higher code speeds is due in part to withdrawal of students from college.

Table 1 presents some illustrative data on the total training time required for students to qualify at four successive levels of code speed under the two conditions described above. Code classes were held one "hour" daily (50 minutes), five days a week, and the practice periods were devoted wholly to receiving. All students were inexperienced in code at the beginning of training and all but three or four were able to master the 10-w.p.m. speed before the course was ended. A few men reached the 15-w.p.m. level and one reached 20 w.p.m., in less than fifty-five hours, but the number is too small to warrant table entries. Three men only failed to qualify at 5 w.p.m. The difference between the means for the two groups is in part due to the fact that, under the second variation of procedure, the slower men were retarded because of the infrequency of runs at the speed which they were seeking to master.

Received February 23, 1944.

Job Specifications in Applied Psychology

Note: The editor will be glad to receive job descriptions and specifications in the fields of applied psychology. These will be published from time to time to provide information for students, applied psychologists, employers, and employment agencies.

Personnel Technician (Tests and Measurements) *

Introduction. This series includes all classes of positions, the duties of which are to construct and administer or supervise the construction and administration of psychological measures, including aptitude tests, intelligence tests, achievement and trade tests, questionnaires, rating scales, etc., to be used as aids in the selection, placement, training, and promotion of civilian and military personnel.

It is believed that the following statements of standards for the positions of personnel technicians (tests and measurements) cover the majority of operating positions in the War Department field service. Positions of higher grade are found on the staff level. These staff positions involve the formulation of policy and procedures to be used in the construction and administration of psychological measures throughout the War Department.

Two general types of positions operating in the War Department field installations are representative of classes properly allocable to this series: 1. Positions involving the construction, validation, and standardization of psychological measures. 2. Positions involving the administration, scoring of psychological measures, and the interpretation and application of the results of these measures to the placement, training, and promotion of civilian and military personnel.

Positions in the Personnel Technician Series may be allocated from grades P-1 through P-5 depending upon the following factors which are interrelated and should not be considered independently:

1. For positions involving the construction of psychological measures.

a. Degree of instruction and guidance given incumbent by superiors.

b. Presence and degree of responsibility for technical accuracy of completed projects.

c. Presence and degree of responsibility for planning and direction of program and projects.

---

* These job specifications were prepared by the Salary and Wage Administration Branch, Civilian Personnel Division, Office, Secretary of War, for use in classifying civilian positions in War Department field establishments.

d. Responsibility for training other personnel technicians of lower grade.

e. Amount of psychological knowledge required in the performance of assignments.

2. For positions involving the administration of psychological measures.

a. Degree of instruction and guidance concerning technical procedures given incumbent by superiors.

b. Presence and degree of responsibility for recommendations concerning placement, training, and promotion of personnel tested.

c. Degree of interpretation and application of test results.

d. Presence and degree of individual accountability for accuracy of results.

e. Presence and degree of responsibility for planning and direction of program.

The following definitions may aid in the complete understanding of these standards:

Aptitude test. One which predicts degree of success or failure in some field or activity.

Achievement test. One which determines level of skill or range of information which an individual has acquired.

Objective test. Short item tests in which there is little disagreement among those who know the subject matter concerning correctness of answer.

General ability. General intelligence or classification test—a group or battery of tests such as vocabulary, general information, arithmetical reasoning, and spatial relations which measure the grade or level of the individual's mental ability.

Validity. The degree to which a test measures the traits or trait it purports to measure.

Reliability. The degree to which a test produces the same results or score upon retest.

Norm. The distribution of scores obtained from a representative sample of the population for which the test is designed.

Process of standardization. Involves the determination of the validity and reliability and the establishment of norms for psychological tests.

Skill. A developed or acquired ability to perform a task, group of tasks, trade, or craft.

Test battery. A combination of carefully selected tests designed to determine certain aptitudes, abilities, achievements, or skills.

Performance test. A type of mental test in which the role of language is greatly reduced, the test material consisting of concrete objects or pictures, and the responses consisting of manipulation or assembling of these objects, e.g., mechanical assembly test, work sample.

Correlation technique. A statistical method applied in obtaining the degree of relationship existing between two sets of measures, such as test scores. There are several types of correlation techniques. For example, biserial, tetrachoric, multiple, partial, etc. Each technique utilizes a different statistical formula. The determination of the specific technique to be used depends on the nature of the problem, the data available, and the degree of accuracy and form of results desired.

Central tendency. A measure or score representing the middle of a group, e.g., mean, median, or mode.

Dispersion. The extent to which a group of measures such as test scores vary from central tendency.

Personnel Technician, P-1

The area of responsibility of grade P-1 personnel technicians is limited by specific instructions and continuous technical supervision by professional personnel of higher grades. Incumbents at this level usually serve as trainees in order to become familiar with the techniques and tools necessary to perform more responsible work in the field of psychological test construction and administration. All of their work is reviewed for technical adequacy, accuracy, and conformance to standardized procedure.

Personnel technicians at this grade collect background material concerning the subject, trade, or skill for which the test, rating scale, questionnaire, etc., is to be constructed. Following outlined sources of material, they study technical manuals and other sources of occupational information, and interview foremen, supervisors, and skilled workers in the field analyzing their jobs in order to determine the essential skills and steps involved in the adequate performance of the work. They perform assigned statistical analyses of data from test studies conducted in the field and on tests constructed by other members of that unit as part of the test standardization process. Following defined procedures, they apply simpler statistical techniques such as correlation formulae to test data in order to determine the reliability and validity of tests.

Grade P-1 personnel technicians, following specific instructions concerning test specifications, subject matter to be covered, types of items, weights, time element, etc., construct objective test items for a variety of subjects, trades, or skills. They may edit test items for conformance to the accepted style (choice and arrangement of words) written by other personnel technicians at this grade, following detailed instructions and samples of items.

Personnel technicians in this class, applying their professional educational background and following specific instructions, analyze, evaluate, and abstract pertinent material relating to test construction and testing procedures from scientific and professional journals.

Grade P-1 personnel technicians functioning in a testing unit administer and score a variety of individual and group tests of aptitudes and abilities following specific instructions concerning choice of tests and methods of administration and scoring. While working at this level, they receive instruction and guidance from personnel technicians concerning interpretation of test results and application of these results to the placement and training of personnel.

Personnel Technician, P-2

Grade P-2 personnel technicians work under supervision of personnel technicians of higher grade, receiving instructions concerning the general procedures and techniques to be followed in completing an assigned project or portion of a project, but working out details on own initiative, and selecting a specific technique from several standardized procedures. Grade P-2 personnel technicians, although advanced beyond the preliminary training period, are given training and instruction on the more difficult phases of their assignments. Their work is reviewed for technical adequacy, accuracy, and conformance to standardized procedures.

Personnel technicians at this level select sources to be used and collect and analyze technical background material concerning the subject, trade, or skill for which the test, rating scale, questionnaire, etc., is to be constructed. They conduct interviews with foremen, supervisors, and skilled workers in the field, analyzing their jobs in order to determine the essential skills and steps involved in the adequate performance of the work. Following established criteria, they select significant factors from the information gathered and they convert job information into valid objective test items.

Grade P-2 personnel technicians, following instructions concerning test specifications, general format of test, subject matter to be covered, types and number of items, construct objective test items for a variety of subjects, trades, or skills. They may edit test items written by other personnel technicians at this grade for conformance to the accepted style (choice and arrangement of words).

Personnel technicians at this grade write manuals of directions for examiners and instructions for examinees following format and outline used in previous test manuals.

Personnel technicians at this level perform statistical analyses of data from test studies conducted in the field and on tests constructed by other members of the unit in order to standardize the tests. Specifically, they determine which formulae to apply to an assigned problem after general plan for statistical study has been outlined by personnel technicians of higher grade.

Grade P-2 personnel technicians functioning in a testing unit administer, score, and interpret a variety of individual and group tests of aptitudes and abilities, i.e., achievement and general intelligence. At this level personnel technicians select the appropriate tests to be administered to personnel being considered for placement, training, transfer, or promotion in order to determine present level of skill, aptitude, suitability for employment, and need for training. The tests are selected on the basis of information obtained by the placement interviewer and on the knowledge which the incumbent has of the uses of specific approved tests in relation to success in the various occupations. Personnel technicians at this grade make preliminary recommendations to the personnel technician in charge of a testing unit concerning the acceptability, placement, and training of the individuals tested, after interpreting the test results.

Personnel Technician, P-3

Grade P-3 personnel technicians receive general instructions concerning objectives of projects but work out details of assignments on own initiative. Their work is reviewed for conformance to technical procedures and established objectives. Personnel technicians at this level review and edit test items prepared by lower grade technicians. This review is made in order to determine whether or not the items of the test conform to established principles and patterns of test construction, are written in clear English, are free of objectionable content, and of a sufficient level of difficulty for the group to be tested, but not too difficult to be of value in differentiating the group.

Grade P-3 personnel technicians following an established plan and purpose for a test or a battery of tests construct objective tests covering a variety of subjects, trades, or skills, and write manuals of directions for examiners and instructions for examinees.

Personnel technicians at this grade review job analyses and trade test items prepared by personnel technicians of lower grade in order to determine whether or not the significant factors concerning the job have been covered by the test items. In performing this function they draw upon their knowledge of jobs and trade tests previously constructed and validated.

Grade P-3 personnel technicians make field trips to posts, camps, and stations in conjunction with personnel technicians of higher grade for the purpose of administering individual tests of mechanical and other aptitudes, and obtaining data for the purpose of validating tests by comparing test scores with efficiency ratings and other personnel data.

Personnel technicians at this level are responsible for statistical analysis of test data. This function involves selecting specific statistical techniques that pertain to the problem from several accepted methods.

Personnel technicians at this grade assemble and organize information for a project report indicating procedures used in gathering facts and preparing lists of data obtained, analysis of statistical results, validation and standardization techniques and norms established, or conclusions reached from the study.

Grade P-3 personnel technicians may function as assistant chiefs of placement testing units in installations having a large variety of complex jobs. In this capacity they determine work assignments for personnel technicians in the unit on the basis of knowledge of individual technician's experience and qualifications. They discuss these assignments with personnel technicians of lower grade and make decisions concerning tests to be administered and recommendations to be made in instances where the subordinate personnel technicians cannot make the determination.

Grade P-3 personnel technicians may also function as chiefs of placement testing units in installations having a smaller variety of routine and semiroutine jobs. They are responsible for the development and administration of a testing program for improving the placement and training of employees in order to achieve the maximum utilization of manpower skills available to the installation. In this capacity they guide and direct a group of personnel technicians of lower grade engaged in administering, scoring, and interpreting the results of aptitude, performance, trade information, and general intelligence tests. Personnel technicians at this grade make recommendations to the placement and training branches of the installation's personnel office concerning the acceptability, placement, and training of the individuals tested, on the basis of the interpretation of test results and analysis of abilities, skills, and aptitudes of the individuals.

Personnel Technician, P-4

Grade P-4 personnel technicians function independently with regard to technical procedures, but work in conformance with established policies concerning objectives to be accomplished.

Personnel technicians at this level are assigned special studies requiring a broad background of psychological measurements and test construction techniques, i.e., evaluating tests and rating scales already in use by comparing results of different tests. They study groups used in original validation of tests and rating scales in order to determine what other groups can be tested by these measures. On the basis of these studies they submit recommendations to personnel technicians of higher grade for improvements in specific tests and test construction techniques.

Personnel technicians at this grade have the responsibility for the planning and completion of projects involving the construction of tests and test batteries, questionnaires, rating scales, etc., which measure general ability, aptitude, or achievement to classify individuals on the job, classify them for specialized training, or classify them within training groups, or evaluate the effectiveness of the training program. They are responsible for project reports which include procedures used in gathering facts and preparing tests, data obtained, analysis of statistical results, validation and standardization techniques and norms established, or conclusions reached from the study.

The details of the test construction function include performance or the supervision of the following:

a. Analysis of the job, field, ability, or skill for which the test is to be constructed.

b. Establishment of test specifications including the determination of the type of test to be constructed, the subject matter to be covered, the number of items, weights, time element, etc.

c. Writing of original test items.

d. Organization of the items into tests.

e. Determination of the format of the test.

f. Organization of the tests into batteries.

g. Selection of standardization groups and design of test tryouts.

h. Planning and extension of statistical analyses for determining validity and reliability of the tests, test batteries, etc.

i. Planning and execution of follow-up studies on the tests constructed.

Personnel technicians at this level accompany higher grade personnel technicians on visits to Army installations after requests have been received from such installations for the establishment of personnel procedures requiring the use and application of psychological measures. In conjunction with higher grade personnel technicians they analyze the test needs in terms of the types of jobs found at the installation and the training and skills of the personnel of the installation. The tests may be used as aids in selection, placement, or promotion of personnel and the evaluation of the training programs. In making this analysis, the functions and work flow of the installation are studied, personnel are interviewed, and administrative officials are consulted. After recommendations concerning testing programs have been approved by the senior member of the team, they instruct installation personnel in uses of tests, methods of administration, scoring, and interpretation. They construct new tests when existing ones do not fit the problem at hand or adapt or extend norms of existing measuring devices to cover the installation personnel.

Grade P-4 personnel technicians advise personnel technicians of lower grades concerning statistical procedures to be applied to particular, unprecedented, or unusual problems. They plan statistical techniques and steps to be used in the determination of the validity and reliability of tests. The steps include intercorrelations of a number of variables, measures of central tendency, measures of dispersion, selection of the minimum number of independent variables which will give maximum production of the dependent variable, item analysis by biserial or tetrachoric correlations and application of measures which indicate whether or not differences obtained are significant. When necessary, new statistical procedures are devised to meet special problems.

Grade P-4 personnel technicians function as chiefs of placement testing units in installations having a large variety of complex jobs. They are responsible for the development and direction of a placement testing program in order to achieve the maximum utilization of manpower skills available to the installation. In this capacity they guide and direct a group of personnel technicians of lower grade engaged in administering, scoring, and interpreting the results of aptitude, performance, trade information, general intelligence tests, questionnaires, and rating scales.

Personnel technicians at this grade make recommendations to the placement and training branches of the installation's personnel office concerning the acceptability, placement, and training of the individuals tested on the basis of the interpretation of test results and analysis of the abilities, skills, and aptitudes of the individuals.

Personnel Technician, P-5

Grade P-5 personnel technicians act with independence in planning projects and in the development of methods and procedures for the conduct of these projects. They receive guidance concerning administrative and technical policies and broad objectives of the program. Their work is reviewed for conformance to these policies and objectives.

Personnel technicians at this level may exercise administrative supervision and technical direction over subordinate professional employees, making final decisions with respect to the most difficult and unusual technical procedural problems which arise.

Personnel technicians at this grade visit War Department installations after requests have been received from such installations for the establishment of personnel procedures requiring the use and application of tests in order to increase the effectiveness of induction, placement, promotion, and training of civilian employees in clerical, mechanical, subprofessional, professional, and supervisory positions. In this capacity they analyze and evaluate the total personnel program in relation to the specific problem in order to determine whether or not the application of tests will alleviate the problem. When such study reveals that a testing program is called for, they select the appropriate previously constructed tests, or construct new measuring devices to meet the problem. The performance of this function involves the following:

a. Analysis of jobs, flow of work, functional organizations, and training curricula and standards.

b. Evaluation of qualifications and performance of present incumbents.

c. Conferences and interviews with employees, supervisory and administrative officials, and with the members of the staff of the installation civilian personnel office.

d. Planning and developing of new test forms and procedures.

e. Preparation of manuals of direction for administration and scoring of tests and interpretations of results.

f. Determination of adequate criteria for test validation and devising procedures for measuring individual job productivity, both qualitatively and quantitatively.

g. Experimental tryouts of preliminary test forms, statistical analysis of data obtained, revisions for final test battery, and establishment of norms for group.

h. Training of installation staff in method of administration and scoring of tests and interpretation of results.

i. Conducting of follow-up studies of testing programs instituted in order to recommend revisions in standards or in measuring instruments.

On these visits to field installations, personnel technicians at this level advise and assist management on other phases of personnel administration such as training, placement, and employee relations, applying their psychological knowledge and experience to the specific personnel problems of the installation.

Grade P-5 personnel technicians may plan, organize, and direct the work of a staff of professional employees of lower grades engaged in the construction and statistical validation of a variety of tests and test batteries to be used in the selection, placement, training, and promotion of employees.

Book Reviews

McMurry, Robert N. *Handling personality adjustment in industry.* New York: Harper and Brothers, 1944, pp. xi+297. $2.50.

The author states that this book is written to give business executives insight into the sources and solutions of problems in personnel administration and industrial relations. Although this ambitious goal is not fully attained, the book will give a fresh point of view to any business leader who is progressive enough to read and consider its contents. Perhaps the book's greatest contribution, however, will be to give academic industrial psychologists greater insight into actual industrial conditions.

The material on labor relations is the outstanding section of the book. Several original and provocative ideas regarding labor-management relationships are presented. The author contrasts the ineffective means of solving labor problems with some techniques that have been effective in determining the true causes of strife: the exit interview, grievance committee, employee opinion poll, and the home interview.

For the most part, clinical psychology, psychiatry and psychoanalysis have neglected the industrial field. Industry, in turn, has not recognized the importance of the personality and emotional factors that affect it. The material in this book will help to remedy this situation. Many books on industrial psychology emphasize the measurement of aptitudes and intellectual factors. This book, on the contrary, emphasizes emotional and motivational factors necessary for success on the job.

Much emphasis is placed on the diagnosis of personality disorders but very little attention is given to treatment techniques. No mention is made of non-directive therapy which is being used with good results in several industrial situations. The author points out that many business executives have personality problems themselves, but the suggestion is never made that these disorders might yield to treatment.

Much of the material in the sections on selection and training is valuable although not related to the topic of personality problems in industry. The chapter on "The Home Interview" is presented entirely as a selection procedure, although it would have been more relevant if centered around the theme of personality problems.

The last thirty-three pages of the book constitute a manual to train employment interviewers in the theory of personality. The psychological principles set forth are questionable in places, as on page 259 where we read:

"The emotions, on the other hand, do not need to mature. They function perfectly at birth. The day-old infant experiences rage and fear as intensely as does the adult. In short, emotional reactions do not have to be learned."

Business and academic people interested in human relationships in industry will derive considerable help from this book.

Charles C. Gibbons

Owens-Illinois Glass Company
Toledo, Ohio

Gallup, George. *A guide to public opinion polls.* Princeton: Princeton University Press, 1944, pp. xviii+104. $1.50.

Applied psychologists will welcome this book because it contains just the type of information to which skeptics or the uninformed may be referred when they desire to learn something about this relatively new technique. It deals with twelve general topics such as the function of public opinion polls in a democracy, the cross section, formulating questions, size of the sample, and polling accuracy.

The plan of the book is novel and effective. From the first page to the last, it gives a series of questions together with brief answers. The total number of questions is 80 covering almost every conceivable aspect of polling. The answers, though brief, are straight-forward, factual, and convincing.

In dealing with the question of accuracy in the prediction of elections, Gallup makes out a good case for the steady reduction of error from the notorious 19 per cent error of the *Literary Digest* in 1936 to the one or two or three per cent errors in the prediction of elections in later years. In the opinion of the reviewer, however, this problem is not faced with complete candor. The *constant* errors in the *Literary Digest* fiasco are described at some length but the presence of such *constant* errors in the Gallup 1940 presidential election poll is not mentioned. The reviewer refers to the fact that the 1940 Gallup poll underestimated the Roosevelt vote in state after state. Surprisingly enough, the Gallup poll was unable to overcome this type of error in 1944 when it overestimated Roosevelt's vote in ten states and underestimated it in 38. But this specific defect should not blind one to the general accuracy and the widespread utility of the Gallup poll and of other equally reliable polls. They are important instruments in ascertaining public opinion not only with respect to social and political issues, but also in market research, industrial relations, education, and all other fields where opinions, beliefs, and attitudes must be taken into consideration in dealing with human beings.

Donald G. Paterson

The University of Minnesota

Dvorine, Israel. *Color perception testing charts* (*Vol. 1*); *Color perception training charts* (*Vol. 2*). Baltimore: Published by the author, 1944. $25.00.

These charts are of the pseudo-isochromatic type. There are 60 perception and 70 training plates. Tests for naming colors and for training in color naming are included. The colors employed are based on the subtractive color theory. The author erroneously lists the three primary colors as red, yellow and blue instead of red, green and blue. The perception test is not limited to investigation of red-green and blue-yellow deficiencies but yields information on perceiving both primary and secondary colors. The author considers that some individuals who are listed as color blind are only confused by certain color combinations. These, although classified as color blind, readily improve when given color perception training. It is claimed that these charts will avoid uncertainty and error in measurement of color perception. The only classifications attempted from responses are: normal in naming and perceiving colors, correct naming with defective perception, incorrect naming and normal perception, incorrect naming and defective perception. There is no attempt to designate a person as red-green color blind, etc.

The training charts "educate" or coach people to name colors and to pass pseudo-isochromatic color-blind tests such as for the Navy or the Air Corps. It would seem that *the widespread use of these training tests would tend to nullify the usefulness of all pseudo-isochromatic tests of color vision.*

Miles A. Tinker

University of Minnesota

New Books, Monographs, and Pamphlets

Crime and the human mind. David Abrahamsen. New York: Columbia University Press, 1944. Pp. xiv + 244. $3.00.

Dvorine Color Perception Training Charts. Israel Dvorine, O.D., 2328 Eutaw Place, Baltimore 17, Maryland. 2 volumes. 130 color charts. $25.00.

Absenteeism: management's problem. John B. Fox and Jerome F. Scott. Boston: Division of Research, Harvard Business School, 1943. $1.00.

Conserving marriage and the family. Ernest R. Groves. New York: The Macmillan Company, 1944. Pp. 138. $1.75.

From gods to dictators. Pryns Hopkins. Girard, Kansas: Haldeman-Julius Publications, 1944. Pp. 168. Paper $1.00. Cloth $1.65.

Experiments on the effects of music on factory production. Willard A. Kerr. Applied Psychology Monograph. Stanford University: Stanford University Press, 1944. $1.00.

Psychotherapy in medical practice. Maurice Levine. Chicago: Macmillan Company. Pp. xiv + 320. $3.50.

Teamwork and labor turnover in the aircraft industry of Southern California. Elton Mayo, George F. Lombard and associates. Boston: Division of Research, Harvard Business School, 1944. Pp. 30, 24 charts, 8 tables. $1.00. (Business Research Studies No. 32.)

Learning by exposure to wrong forms in grammar and spelling. John R. McIntosh. New York: Bureau of Publications, Teachers College, Columbia University. Pp. 61. $1.75.

Emotional factors in learning. Lois B. Murphy and Henry Ladd. New York: Columbia University Press, 1944. Pp. x + 410. $3.50.

Women and men. Amran Scheinfeld. New York: Harcourt, Brace & Company, 1944. Pp. 453. $3.50.

Psychodrama in the schools. Nahum E. Shoobs. New York 17: Beacon House, 1944. Psychodrama Monographs, No. 10. Pp. 19. $1.50.

Glossary of technical terms. Calvin P. Stone. Stanford University: Stanford University Press, 1944. Pp. 15. $.25.

A factorial study of perception. L. L. Thurstone. Chicago: University of Chicago Press, 1944. Pp. 148. $2.50.

The place of reading in the elementary school program. Board of Education of the City of New York. Educational Research Bulletin No. 7. Pp. 43.

Journal of Applied Psychology

Vol. 20, No. 3 June, 1945

Studies in Job Evaluation: II. The Adequacy of Abbreviated Point Ratings for Hourly-Paid Jobs in Three Industrial Plants

C. H. Lawshe, Jr.

Division of Education and Applied Psychology, Purdue University

The growing importance of job evaluation as a more objective approach to the stabilization of industrial wage structures has manifested itself in the increasing degree [1] to which psychologists are concerning themselves with the various rating methods now in use. To examine by psychological methodology some of these job rating techniques has been the intent of the authors of this series of papers, the first [2] of which reported the identification of factors or clusters of items which appear to be functioning in one of the most widely used job rating systems.[3] The purpose of the investigation reported in the present paper was to determine the extent to which abbreviations of this system yield the same or comparable results and to examine any differences in terms of practical significance.

Procedure and Results

Method. Job rating data were collected from three plants which use the NEMA system or a slight modification of it. For each plant, intercorrelations between all eleven factors and total points were computed and a correlation matrix was prepared.[4]

---

[1] Herbert Moore. Problems and methods in job evaluation. *J. consult. Psychol.*, 1944, 7, 90-99.

[2] C. H. Lawshe, Jr., and G. A. Satter. Studies in job evaluation. 1. Factor analyses of point ratings for hourly-paid jobs in three industrial plants. *J. appl. Psychol.*, 1944, 28, 189-198.

[3] *Job rating: Definition of the factors used in rating jobs—hourly rated occupations.* Chicago: Industrial Relations Department, National Electrical Manufacturers Association, 1938. Pp. 22.

[4] These matrices together with descriptions of the three plants and other pertinent data are presented in the first of this series of papers, C. H. Lawshe, Jr., and G. A. Satter, *Op. cit.*

The Wherry-Doolittle shrinkage selection method as reported by Stead, Shartle; *et al*[5] was applied and the first three items were identified in each plant. The multiple R's are presented in Table 1.

Table 1

Correlation Coefficients between Ratings on Selected Items and Total Point Ratings

	Plant		
Selected Rating Scale Items	A	B	C
Experience (or Learning time)	.96	.93	.86
Experience (or Learning time) plus Hazards	.97	—	.91
Experience plus Initiative	—	.95	—
Experience (or Learning time) plus Hazards plus Education	.98	—	—
Experience plus Hazards plus Initiative	—	—	.93
Experience plus Initiative plus Responsibility for the Safety of Others	—	.96	—

Items Identified. As is shown in Table 1, the "experience or learning time" item is the single variable which correlates highest with "total points" in each of the three plants, the coefficients for plants A, B, and C being .96, .93, and .86 respectively. In plant A, when "hazards" is added, the multiple correlation becomes .97 and when "education" is added it becomes .98. In plant B, when "initiative" is added the correlation is increased to .95 and when "responsibility for the safety of others" is added, the value is .96. For plant C, the multiple correlations are increased to .91 and .93 with the subsequent inclusion of "hazards" and "initiative."

It should be pointed out that in none of the three plants did the R start to shrink when the third variable was added. However, the high value of the correlations, plus the fact that the increment resulting from the addition of the third variable is so small, makes further application of the technique seem unnecessary. The difference between these increments added to the R's by the third selected items as compared to the increment that would have been added by other items is so small that considerable sampling error could be present. For example, in Plant B, when the correlations are carried to a third decimal place, the addition of "responsibility for the safety of others" to "experience" and "initiative" increases the R from .953 to .962, an increment of .009. Other items would have increased the obtained R by perhaps .007 or .008. It seems, then that too much importance should not be attached to the particular item that was added last. In spite of this fact, however, Table 1 shows a certain consistency from plant to plant in the particular items that were selected. In the three variables selected, "experience or learning time" appears in all plants, always first, while "hazards" and "initiative" each appear in two of the plants.

---

[5] William H. Stead, Carroll L. Shartle; *et al.* *Occupational counseling techniques*, pp. 245–252, New York: The American Book Company, 1940.

Accuracy of Prediction. Table 2 lists the standard errors of estimate for predicting the total point rating in each plant from one, two, and three items in the rating scale. For example, the standard error of estimating the total point rating for a particular job from the best three items is 11.4 in Plant A, 13.0 in Plant B, and 3.8 in Plant C. In other words, in Plant A, the estimates for approximately two-thirds of the jobs are within 11.4 of the total point rating based on all eleven factors. The percentage figures in Table 2 indicate the proportional size of the errors in terms of the standard deviations of the several distributions.

Table 2

Standard Errors of Estimate for Predicting Total Point Ratings from Selected Scale Items

	Plant A		Plant B		Plant C	
Selected Items	$\sigma_{\text{est.}}$	%	$\sigma_{\text{est.}}$	%	$\sigma_{\text{est.}}$	%
Best Single Item	17.4	30	17.0	37	5.4	51
Best Two Items	13.7	24	14.4	30	4.5	42
Best Three Items	11.4	10	13.0	27	3.8	36

Practical Implications of Grade Placement

Application of Abbreviated Scale. In the plants studied, how many jobs would actually be shifted insofar as rates of pay are concerned if an abbreviated scale were used? Is the standard error of estimate of 11.4 in Plant A practically significant? This question can best be answered through an analysis of the changes that would actually occur if only three items were used.

Prediction Formula. Plant A has been selected as an example. Using the data from this plant, the regression equation for predicting total points from "experience or learning time," "hazards," and "education" was found to be:

$$X_{TP} = 30.4 + 1.4_{\text{Exp.}} + 5.4_{\text{Haz.}} + 2.0_{\text{Educ.}}$$

Point ratings on "experience or learning time," "hazards" and "education" were substituted in the formula for each of the 247 jobs in the plant to obtain the computed ratings. These computed values are shown plotted against the total point ratings (eleven items) in the scattergram (representing the previously mentioned R of .98) in Figure 1. Superimposed over the scattergram are eleven shaded areas, each representing a different labor grade. For example, jobs which "rate" from 144 points to 165 points are in the second labor grade. Any job which falls inside a shaded area is placed in the same labor grade by both the original scale and the abbreviated scale, and any job *which falls in an unshaded area* would be displaced one or more labor grades by the abbreviated scale.

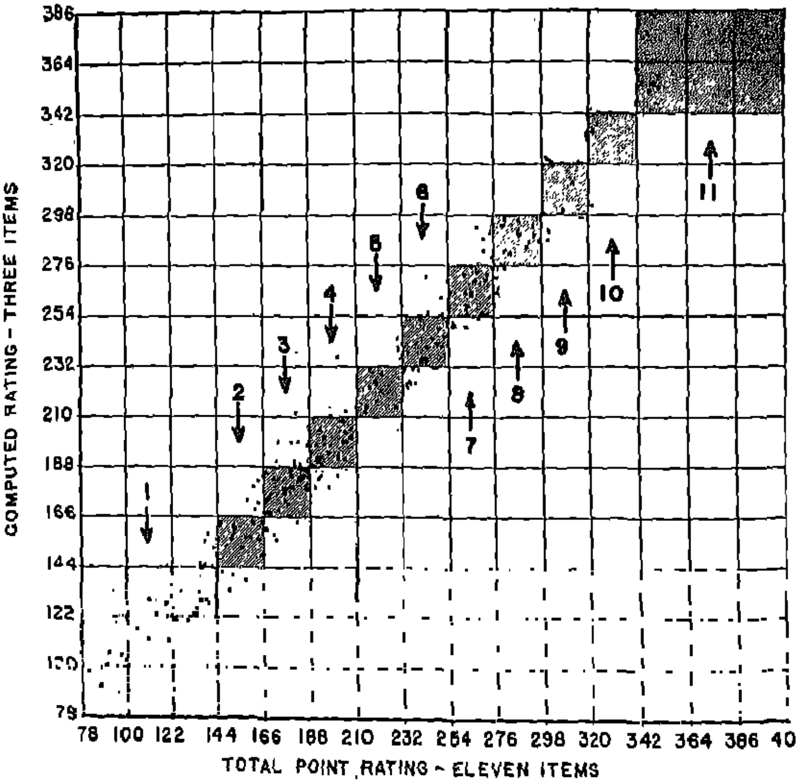


FIG. 1. Graph showing ratings computed from three scale items plotted against total point ratings (all eleven items) for 247 jobs in Plant A. The eleven shaded areas define the labor grades designated by the numbered arrows.

Labor Grade Displacement. Table 3 shows that of the 247 jobs in this particular plant, 153 or 62% would remain in the same labor grade, 92 or 37.2% would be displaced by one labor grade, while only 2 jobs or 0.8% would be displaced two labor grades. Table 3 also shows the number of jobs deviating by varying numbers of points, classified as "same labor grade," "displaced one labor grade," and "displaced two

Table 3

Discrepancies between Total Point Ratings (Eleven Items) and Ratings Computed from Three Items for Plant A

Points of Deviation	No. of Jobs by Labor Grade Displacement: Same Labor Grade	Displaced One Labor Grade	Displaced Two Labor Grades	All Jobs
0–4	68	9		77
5–9	48	27		75
10–14	20	28		48
15–19	11	21		32
20–24	6	2		8
25–29		4		4
30–34		1	1	2
35–39			1	1
Totals	153	92	2	247

labor grades." This table and the data from which it was prepared reveal that only seven jobs deviate from their original placement by more than 22 points, the range of most of the labor grades. The fact that only seven jobs have point differences greater than the difference between the highest rated job and the lowest rated job in any given grade, is additional evidence of the comparability of the two systems.

Wage Structure Considerations

Range of Rates. Examination of the rate schedule (*Table* 4) in Plant A tends to minimize the practical importance of such differences as would

Table 4

Rate Schedule For Plant A

Labor Grade	Point Range	Rates: Starting	One Month	Two Months	Maximum
1	Up to 144	.65	70	.75	.81
2	145 to 166	.65	70	.75	.87
3	167 to 188	.70	75	.80	.93
4	189 to 210	.75	80	.85	.99
5	211 to 232	.80	85	.90	1.05
6	233 to 254	.85	90	.95	1.11
7	255 to 276	.90	95	1.00	1.17
8	277 to 298	.95	100	1.05	1.23
9	299 to 320	1.00	105	1.10	1.29
10	321 to 342	1.05	110	1.15	1.35
11	343 and up	1.10	115	1.20	1.41

exist between the application of the original scale and the abbreviated scale. The rate of $1.05 per hour, for example, is the maximum rate for jobs in labor grade five and is earned by employees on some jobs which are evaluated as low as 211 points. On the other hand, $1.05 is the starting rate for labor grade ten and is paid to some employees on jobs that are evaluated as high as 342 points. This rate, then of $1.05, may be paid at one time or another to employees on jobs ranging through six different labor grades and evaluated at from 211 to 341 points. To be sure, a progression system is used and varying amounts of time are spent on jobs in these several labor grades when the $1.05 rate is paid. However, the fact that a particular rate does occur in six consecutive labor grades tends to minimize the practical significance of a few points of difference on the scale. Figure 2 shows that nearly every job that falls within the 211 to 342 point range on one scale also falls within the same range on the other.

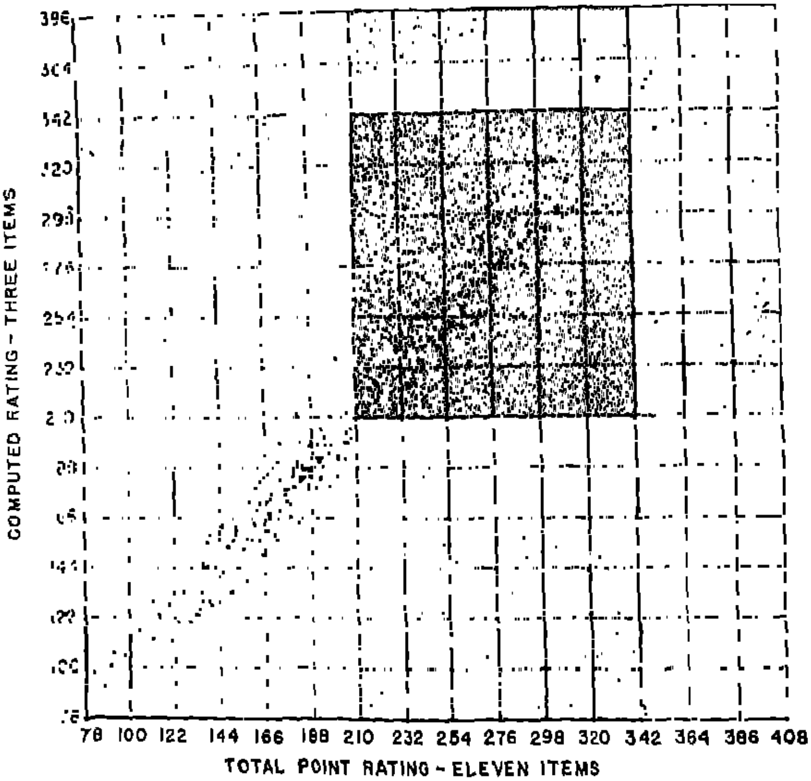


FIG. 2. Graph showing point range of jobs which at one time or another carry a rate of $1.05. Note that only seven jobs fall in one shaded band but not in the other.

The Reliability of Point Ratings

Practical Difficulties. The reliabilities of the original ratings in the three plants investigated are not known just as they are not known in any plant. It is impractical if not impossible to obtain re-ratings of jobs by a second jury of equal competence or of equal familiarity with the jobs. For this reason, it is impossible to determine whether or not the correlation between the complete scale and the abbreviated scale is as high as the reliability of the scale itself. Needless to say the reliability of human judgments rarely attains the magnitude of the correlations presented in Table 1, whether they be judgments of personality traits or of physical phenomena.

Table 5 sheds some light on the reliability problem. In a fourth plant, the same jury of supervisors and analysts rated a group of five jobs on March 16, rerated them on March 22, and again rated them on April 7.

Table 5

Point Ratings and Corresponding Rates for Five Jobs Rated at Three Different Times

	First Rating		Second Rating		Third Rating		Maximum Change	
Job	Points	Rate	Points	Rate	Points	Rate	Points	Rate
A	435	.95	395	.91	335	.85	100	.10
B	405	.92	330	.84	295	.81	110	.11
C	315	.83	260	.78	245	.77	70	.06
D	355	.87	370	.88	380	.89	25	.02
E	330	.84	380	.89	380	.89	50	.05
					Mean Change		71	.05

The table shows that the fluctuations in point ratings ranged from 25 to 110 points with an average change of 71 points, while the corresponding rates that would be paid for jobs with these point values showed changes ranging $.02 to $.10 per hour with an average change for the five jobs of $.05 per hour. Perhaps the jobs changed; perhaps the analysts accumulated additional information for the subsequent ratings; perhaps the analysts improved; or perhaps they got tired. Whatever the reason, the point values did fluctuate. The fact of these fluctuations further minimizes the practical importance of a few points in this particular job rating plan. That there is high agreement between the original system as it functions in the three plants and the abbreviated system is a fact. Whatever the amount of deviation, there has been no intention to imply that either the original scale or its abbreviation is the criterion against which to measure the other. Such deviations from perfect reliability as are almost certain to exist force the conclusion that the complete scale and an abbreviation of it yield results which in terms of practical operation are almost identical.

Summary and Conclusions

Job rating data from three different plants were subjected to the Wherry-Doolittle selection technique following the intercorrelation of points awarded on each item. The three items yielding the highest R with the total point rating from eleven items were identified for each plant and a more detailed study of Plant A was made. The following conclusions are supported:

1. In each of the three plants, the "experience or learning time" item in the scale correlated highest with the total point rating arrived at from eleven items, the r's being .96, .93, and .86.
2. The R for the combination of the three optimum scale items was found to be .98, .96, and .93 for plants A, B, and C, respectively.
3. The items, "hazards" and "initiative" each appeared twice in the optimum three items in the plants studied.
4. If the three item abbreviated scale were employed in Plant A, 62% of the jobs would remain in the same labor grade, 37.2% would be displaced one labor grade, and 0.8% would be displaced two labor grades.
5. Such deviations as exist between point ratings assigned by the original scale and by the abbreviated scale seem practically unimportant in terms of the magnitude of the range within any given labor grade, the flexibility of the plant wage schedule itself, and the probable unreliability of the ratings.
6. A simplified scale consisting of three or four items would probably yield results that are practically identical with those obtained by a more complex system and would greatly reduce the time consumed by the rating activity.

Received June 30, 1944.

The Value of Aptitude Tests for Supervisory Workers in the Aircraft Engine and Propeller Industries

John T. Shuman

Williamsport Technical Institute, Williamsport, Pennsylvania

What the future holds for any company depends in a large measure on the ability, vision, and leadership of those in supervisory and executive positions. Management today is quite generally free from too great dependence on rule-of-thumb and the shrewd-guess method of operating. Time consuming engineering efforts bring new designs into being, provide the tools and other facilities required for manufacture of an article. Yet the problems involved in the selection and training of men as supervisors have received inadequate attention. Generally industrial managers would not countenance such thoughtless procedures in the handling of raw materials as are sometimes used in the selection and development of supervisors. This part of the investigation, however, is confined to only one phase of this problem—that of reporting the results obtained in comparing certain test scores with the job success of the supervisors studied.

In their present state tests are not wholly adequate for predicting supervisory success. However, tests and rating scales do provide good and effective bases from which to start, or checks by which to gauge decisions. Supervisory and executive ability arise from the interaction of many different qualities and abilities, and no single test will measure entirely the many qualities and abilities involved.

Method

The foremen, assistant foremen, set-up men, and group leaders as the case may be, were tested in small groups. These men were already working as supervisors when given the tests; hence, it is safe to assume that some form of natural selection had already taken place in most instances.

The rating sheet illustrated here was used to secure a fairly objective rating on the job success of each supervisor tested. These ratings were made by superiors of the supervisors studied. No more objective measures of job success were available which would have been even relatively free from possible distortion by factors beyond the control of the supervisor. It is true, of course, that the study therefore is valid to the extent that the ratings represent a true picture of the job success of these individuals. Relatively few ratings of "poor" were secured, probably due to the reluctance of superiors to so rate the men. For this reason, many of the comparisons will be made between the "excellent" and "average" ratings. In other words, the percentage of excellent ratings will be used to determine the efficacy of the tests. This procedure is justified for several reasons: 1. The small proportion of poor ratings. 2. The probable operation of the so called "halo effect" among the average ratings. 3. The excellent foreman is the one in whom we are most interested.

Experience in giving these tests to groups of foremen indicates that giving tests to groups of older employes is not too desirable a practice if it is at all possible to test all employes at the time they are hired. Since a man or woman applying for a position is usually willing to take a test, no problem of morale is involved because of the possibility of imminent decisions affecting the man's job in the near future.

Rating Sheet Used in Rating Supervisors

Date——————————

Name——————————

Position——————————

Rated by——————————

1. Production: Consider whether work generally moves through this department on schedule, scrap and rework.	Above Average Good	Average	Below Average Poor
2. Handling workers: Consider discipline; extent to which this man has difficulty with his men; attitude of men toward company policies, etc.	Above Average Good	Average	Below Average Poor
3. Condition and maintenance of department: Good housekeeping, safety, condition of machines.	Above Average Good	Average	Below Average Poor
4. General: In general and from all aspects how would you rate.	Above Average Good	Average	Below Average Poor

Table 1 presents a summary of recommended minimum critical scores and the percentage of improvement effected in the selection of foremen at the three manufacturing plants studied.

Table 1

Summary of Recommended Minimum Critical Scores and Improvement Effected in Selection of Foremen, All Plants

	Minimum Critical Score			Per Cent Improvement in Excellent Ratings		
Test	Lycoming ($N=99$)	American Propeller ($N=80$)	Spencer Heater ($N=24$)	Lycoming ($N=99$)	American Propeller ($N=89$)	Spencer Heater ($N=24$)
Otis Q.S. Test of Mental Ability Beta, A	33	30	32	9	5.4	42
Minnesota Paper Form Board, AA	24	34	18	8	5.4	21
Bennett Test of Mechanical Comprehension, AA	30	30	25	10	8	12

The per cent improvement effected in the selection of Foremen was greatest at the Spencer Heater Plant and least at the American Propeller Plant with the results at the Lycoming Plant between the two. It is significant, however, that an improvement in selection would be possible at each of the three plants with all of the tests used.

Table 2 summarizes the results with group leaders and job-setters at two of the plants. The third plant, Spencer Division, had no supervisors in this category. The percentage of improvement in selection possible at both the plants represented in Table 2 is rather high. The results secured

Table 2

Summary of Recommended Minimum Critical Scores and the Improvement Effected in Selection of Job-Setters and Group Leaders, All Plants

	Minimum Critical Score		Per Cent Improvement in Excellent Ratings	
Test	Lycoming ($N=25$)	American Propeller ($N=60$)	Lycoming ($N=25$)	American Propeller ($N=60$)
Otis Q.S. Test of Mental Ability Beta, A	34	34	17	24
Minnesota Paper Form Board, Revised, AA	30	24	30	12
Bennett Test of Mechanical Comprehension, AA	36	27	47	14

here with the group leaders in the American Propeller Corporation are much higher and more positive then those secured in the same company for foremen and reported in Table 1.

As indicated in Table 3, the use of these three tests would improve generally the selection of excellent supervisors by 15 to 20 per cent.

Table 3

Average Improvement Effected in Selection of Excellent Supervisors,* All Plants, All Supervisors, $N = 207$

Test	Mean Improvement in Selection Excellent Supv. at Minimum Critical Scores	Mean Improvement in Selection of Excellent Supv. at Q_1
Bennett Test of Mechanical Comprehension, AA	18%	20%
Otis Q.S. Test of Mental Ability Beta, A	19%	17%
Minnesota Paper Form Board, Revised, AA	15%	17%

* Includes all supervisors: foremen, assistant foremen, group leaders, and job setters.

In Table 4 statistically significant correlations were secured on the Otis test in the Lycoming and Spencer Heater plants; on the Bennett test in the Lycoming and Spencer plants; and on the Minnesota Paper Form Board in only the Lycoming plant. No statistically significant correla-

Table 4

Summary of Correlations between Job Ratings and Test Results, Foremen and Assistant Forement, All Plants

Tests	Job Rating Tests with Raw Scores		
	Lycoming Division ($N = 99$)	American Propeller Corporation ($N = 80$)	Spencer Heater Division ($N = 24$)
Otis Q.S. Test of Mental Ability Beta Test, A	.39 ± .07	.05 ± .09	.66 ± .12
Minnesota Paper Form Board, Revised, AA	.47 ± .07	.08 ± .09	.14 ± .04
Bennett Test of Mechanical Comprehension, AA	.465 ± .07	.02 ± .09	.65 ± .12

tions were secured on any one of the three tests at the American Propeller Corporation Plant.

Table 5 summarizes the correlations obtained on the group leaders and job setters. The correlations with but one exception are significant. This one exception, an r of .37 ± .10, on the Bennett Test with the American Propeller Corporation group leaders, approximates a significant correlation. Again, the results with this group of American Propeller

Table 5

Summary of Correlations between Job Ratings and Test Results, Group Leaders and Job Setters, All Plants

Test	r_{bis} Job Rating with Test Scores	
	Lycoming Division ($N=25$)	American Propeller ($N=60$)
Otis Q.S. Test of Mental Ability Beta A	.46 ± .14	.54 ± .08
Minnesota Paper Form Board, Revised, AA	.59 ± .13	.39 ± .09
Bennett Test Mechanical Comprehension, AA	.73 ± .10	.37 ± .10

Table 6

Mean Correlations between Job Ratings and Test Scores, All Supervisors, All Plants

Test	Mean Biserial r's Supervisors All Plants ($N=297$)	Mean Biserial r's Supervisors All Plants Except Amer. Prop. Foremen * ($N=208$)
Bennett Test of Mechanical Comprehension, AA	.46 ± .04	.55 ± .04
Otis Q.S. Test of Mental Ability Beta, A	.42 ± .04	.51 ± .045
Minnesota Paper Form Board, Revised, AA	.33 ± .04	.39 ± .05

* *The results secured with the Foremen, American Propeller Corporation, did not* coincide with those secured for the other four groups of supervisors including the Group Leaders from the same Company. Since these results were so much out of line, the author is taking the liberty of using a mean in this column which excludes this group.

Corporation supervisors are much more positive than the results reported for the foremen of the same company and reported in Table 4.

Averaging the *r*'s secured with all supervisory groups yields the results shown in Table 6. The Bennett Test proved to be the most effective, the Otis Test next most effective, and the Minnesota Test the least effective of the three.

Conclusions

1. Job-success on supervisory work was found to be related positively and significantly to the test scores in these three dissimilar industrial plants engaged in some phase of metal working. It is significant that each of these plants was entirely different from the others in the product manufactured and the character of work performed in the plant. It is reasonable to conclude that supervisory work in these different plants has some factors, skills, or aptitudes in common measured by the tests.
2. The promotability of minor supervisors such as group leaders and job setters was found to be related positively and significantly with the test scores.
3. The means obtained by the foremen and group leaders on the tests at Lycoming were below those obtained by workmen in the more skilled job categories in the factory. This is probably due to the effect of additional experience and service of this group; that is, years of experience on the job count heavily in recommendations for promotion. Furthermore, the tendency of individuals of superior ability to leave routine jobs would have a tendency to drain off many capable individuals before they had acquired a sufficient background of experience to be given the responsibility of supervisory work.
4. It is also logical that tests might well be utilized to survey a plant in order to determine whether its supervisory force measures up to recognized standards. In other words, the results indicate that there are levels below which the supervisory force of a plant should not fall. In this way, the proper tests would become one objective measure of a supervisory force, almost wholly dissociated from experience.

Received March 15, 1944.

Relationship between Interests and Abilities: A Study of the Strong Vocational Interest Blank and the Zyve Scientific Aptitude Test

Louis Long

Student Personnel Bureau, College of the City of New York

If the vocational counselor is to do an adequate job he must constantly seek to discover relationships between the various techniques that he uses. The present study arose from speculation about the relationship between interests and ability as determined by two standardized techniques: The Strong Vocational Interest Blank (6) and the Zyve Scientific Aptitude Test (8). Since most studies have reported a slight positive relationship between interests and abilities the question arose as to whether or not students who rated high on the scientific scales of the interest blank would do better on the Scientific Aptitude Test than students who rated low. Casual inspection of scores as they were discussed in interviews suggested that there was a positive relationship and the results of this study confirm this impression.

Such a relationship is certainly to be expected if the two techniques measure what they purport to measure. In this connection it should be mentioned that there is some question as to what the Scientific Aptitude Test measures (2). Only a slight relationship has been found between the scores on this test and college grades (1, 4). Nor does there seem to be any great amount of communality between the Scientific Aptitude Test and tests of general intelligence (1). If this test is measuring scientific aptitude it is to be expected that students who rate high on the Strong scales for engineers or physicists would make, on the average, higher scores on the Scientific Aptitude Test than do students who rate low on the scales for these occupations. If the Scientific Aptitude Test is measuring scientific ability it can also be anticipated that there is a greater degree of relationship between scores on this test and high ratings on the "scientific scales" of the Strong questionnaire than between scores on this test and high ratings on the "nonscientific scales" of the Strong questionnaire. If such hypotheses are verified by an analysis of the scores on the two instruments it can then be inferred that the Scientific Aptitude Test is measuring some phase of ability that separates students who have interests similar to the scientist from those who do not have interests similar to the scientist. Consequently a more positive relationship between the Scientific

Aptitude Test and the scientific scale of the Strong than between the Scientific Aptitude Test and any other scale of the Strong would suggest that the two instruments should be used to supplement each other in guidance work.

Scores Used in the Statistical Analysis

Strong has published norms not only for the individual occupations but also for groups of occupations (6). At the Student Personnel Bureau of the College of the City of New York it has become standard practice first to score the Strong blank against group occupational keys. If a break down of any group is desirable this can be done later. This procedure greatly reduces the scoring time and on the basis of the reported correlations between the group occupational keys and the individual occupational keys this short cut seems justifiable (5, 6). The occupational groups and the occupations included in each are listed in Table 1. Since this study is primarily concerned with the Strong questionnaire as a tool for use in guidance work, the statistical treatment will be oriented around the letter ratings that Strong uses to indicate the degree of common interest between the individual and a particular occupational group. The raw scores could have been used, but it was thought that the results would be more typical if the letter ratings were employed. In several places throughout the report a division has been made between students obtain-

Table 1

List of Occupations Included in each of the Six Occupational Groups of the Strong Vocational Interest Blank

- Group 1. Technical Non-mathematics
    - Artist
    - Architect
    - Psychologist
    - Physician
    - Dentist
- Group 2. Technical Mathematics
    - Mathematician
    - Engineer
    - Chemist
    - Physicist
- Group 3. Welfare
    - Y. M. C. A. Physical Director
    - Y. M. C. A. General Secretary
    - Personnel Worker
    - Social Science Teacher
    - City School Superintendent
    - Minister
- Group 4. Business Detail
    - Accountant
    - Office Worker
    - Purchasing Agent
    - Banker
- Group 5. Business Contact
    - Sales Manager
    - Real Estate Salesman
    - Life Insurance Salesman
- Group 6. Verbal
    - Advertising Man
    - Lawyer
    - Author
    - Journalist

ing an A or B+ rating and those obtaining a B, B−, C+, or C rating. This dichotomy is arbitrary, but such a division is often made in counseling on the basis of Strong's suggestion that "a person should consider seriously those occupations in which he receives A or B+ ratings before entering some other occupation" (6).

The total score of the Scientific Aptitude Test has been used throughout this report. The scoring procedure recommended by Zyve (8) was followed.

Subjects and Test Scores

Scores on both the Strong Vocational Interest Blank and the Scientific Aptitude Test were available for 200 students of the College of the City of New York. Each student had sought advice from some staff member of the College's Student Personnel Bureau and at the suggestion of the latter had taken both the Strong and the Zyve. Scores on the Thurstone A. C. E. Psychological Examination were also available for these students.

Results

The average score on the Scientific Aptitude Test for each letter rating of the six occupational groups will be found in Table 2. For example, the average score on the Scientific Aptitude Test for students obtaining an A rating on the Strong Technical Non-mathematics group was 107.4. A positive trend will be noted in the case of the Technical Non-mathematics and Technical Mathematics groups; a negative trend is found in the case of the Business Contact and Verbal groups; but no definite trend is apparent in the two remaining groups: Welfare and Business Detail. To form some idea of the significance of these trends the six steps on the Strong scale were consolidated into two steps: ratings of A and

Table 2

Relationship between Scores on Zyve Scientific Aptitude Test and Ratings on Strong Vocational Interest Blank

Occupational Groups of Strong	Average Score on Zyve Test According to Categories of Strong					
	A	B+	B	B−	C+	C
Tech. Non-Math.	107.4	96.5	93.5	93.0	100.0	85.0*
Tech. Math.	111.4	106.3	88.2	90.8	91.9	87.4
Welfare	96.7	101.9	98.9	105.2	92.0	119.8*
Business Detail	84.1	103.4	93.9	97.2	97.0	106.6
Business Contact	88.8	82.4	94.8	99.7	105.2	107.6
Verbal	94.5	97.6	100.0	101.9	104.9	115.4*

* Average based on less than 10 cases.

Table 3

Reliability of the Difference between Average Scores on Zyve Scientific Aptitude Test for High and Low Ratings on Strong Vocational Interest Blank

Occupational Groups of Strong	Subjects with Ratings of B+ or more on Strong		Subjects with Ratings of B or less on Strong		P-value of the Difference between Averages
	N	Average Score on Zyve Test	N	Average Score on Zyve Test	
Tech. Non-math.	110	104.2	90	93.6	<.01
Tech. Math.	97	109.9	103	89.6	<.001
Welfare	108	98.7	92	100.4	>.60
Business Detail	24	94.5	176	100.1	>.30
Business Contact	34	85.4	166	102.4	<.001
Verbal	84	96.0	116	102.0	.10

B+ and ratings of B or less. The average score on the Scientific Aptitude Test for students falling into these two categories will be found in Table 3. The reliability of the difference between the averages was determined by using Student's t-Test (3). The P-values presented in Table 3 indicate that in three of the six cases such a difference between the averages would be expected to occur by chance less than five times in 100. Using this as a criterion of a significant difference we can say that the following trend is significant: students scoring high on the Scientific Aptitude Test rate higher on the Strong Technical Non-mathematics and Technical Mathematics groups, but lower on the Business Contact group than do those scoring low on the Scientific Aptitude Test.

Table 4

Bi-serial Correlations between Scores on the Scientific Aptitude Test and Ratings on Strong Vocational Interest Blank (B+ or more versus B or less)

Occupational Groups of Strong	Bi-serial Correlation
Tech. Non-math.	0.26
Tech. Math.	0.50
Welfare	−0.04
Business Detail	−0.12
Business Contact	−0.37
Verbal	−0.14

If the Scientific Aptitude Test is positively related to any occupational group of the Strong scale it would be expected that the Technical Mathematics group would be the most likely one to show this relationship. The second most likely group would be the Technical Non-mathematics. Positive relationships were found in both instances. The extent of the relationship is, however, greater for the Technical Mathematics group since the P-value in one case is 0.001 and only 0.01 in the other. This difference can also be brought out by calculating a bi-serial correlation between scores on the Scientific Aptitude Test and the categories of B+ or more versus B or less on the Strong scale. When this is done a bi-serial r of 0.50 is obtained for the Technical Mathematics group and one of 0.26 for the Technical Non-mathematics. (The correlations for the other occupational groups will be found in Table 4.)

The negative or chance relationship between the ratings on the four other occupational groups and scores on the Scientific Aptitude Test strengthens the above finding since there is no reason to expect a direct correspondence between the two techniques in these cases.

In an effort to determine whether the students with high scores on the Technical Mathematics group of the Strong scale were just superior students or "high scorers" the same analysis was applied to the total score made on the Thurstone A. C. E. Psychological Examination. Of the 200 students in the preceding comparisons, 95 took the 1938, 1939, or 1940 edition of the Thurstone test. The scores on the 1938 and 1940 editions were converted into 1939 equivalent scores by using the table supplied by Thurstone (7). The average scores on the Thurstone test for students obtaining an A, B+, B, B−, C+, or C rating for the six occupational groups of the Strong scale will be found in Table 5. The absence of definite relationships between scores on the Strong scale and scores on the Thurstone test is clearly evident, except in the case of the Verbal group. When the average Thurstone score for students rating high on the Strong scales was compared with the average Thurstone score for students rating low on the Strong scales a significant difference was found only in the case of the Verbal group (Table 6). It seems logical to expect a relationship between the ratings on the Verbal group of the Strong scale and the scores on the Thurstone test, since the latter is so highly verbal.

Table 5

Relationship between Scores on Thurstone A. C. E. Psychological Examination (1939) and Ratings on Strong Vocational Interest Blank

Occupational Groups of *Strong*	Average Score on Thurstone According to Categories of Strong					
	A	B+	B	B−	C+	C
Tech. Non-math.	124.0	111.4*	124.7	114.3	135.3*	132.0*
Tech. Math.	121.3	129.8	120.0	121.8	132.9*	120.6*
Welfare	125.0	127.4	118.9	127.5*	122.0*	116.6*
Business Detail	116.1*	110.7*	126.1	118.7	128.1	124.4
Business Contact	145.0*	100.8*	122.0	127.1	121.1	119.3
Verbal	120.8	121.7	120.0	120.5	111.7*	116.0*

* Average based on less than 10 cases.

Table 6

Reliability of the Difference between Average Scores on Thurstone A. C. E. Psychological Examination (1939) for High and Low Ratings on Strong Vocational Interest Blank

Occupational Groups of Strong	Subjects with Ratings of B+ *or more on Strong*		Subjects with Ratings of B *or less on Strong*		*P*-value of the Difference between Averages
	N	Average Score on Thurstone	*N*	Average Score on Thurstone	
Tech. Non-math.	49	121.7	46	124.8	>.40
Tech. Math.	46	123.6	50	122.8	>.80
Welfare	49	125.8	46	120.4	>.10
Business Detail	10	114.5	85	124.2	>.10
Business Contact	15	126.2	80	122.0	>.50
Verbal	40	128.5	55	119.3	<.05

Summary

The relationship between interests and scientific ability as measured by the Strong Vocational Blank and the Zyve Scientific Aptitude Test was investigated. Use of the occupational group scales on the Strong disclosed that in three of six groups there was a reliable difference between the average score on the Scientific Aptitude Test for students rating high on the interest questionnaire and for those rating low. The findings indicate that students scoring high on the Zyve Test rate higher on the Strong Technical Non-mathematics and Technical Mathematics groups, but lower on the Business Contact group than do those scoring low on the Zyve Test. In order to rule out the possibility that these results were due to the selection of superior students the same analysis was applied when scores on the Thurstone A. C. E. Psychological Examination (1939) were substituted for scores on the Zyve Scientific Aptitude Test. No definite relationship between scores on the Thurstone test and ratings on the Strong scales was found except in the case of the Verbal group.

The results indicate that the Zyve Scientific Aptitude Test is measuring some phase of ability that separates students having interests similar to those found among the occupational groups included in the Technical Mathematics and the Technical Non-mathematics groups of the Strong from those students who do not have interests similar to those found in the above occupational groups. From the common sense point of view the agreement between the two measurements is to be expected. The Zyve Test deals largely with problems involving mathematics and principles of physics. Even when the items are intended to measure general abilities (such as reasoning, generalizing, and suspending judgment) the content of the item is usually drawn from the physical science field. Similarly, inspection of the keys of the Strong Technical Mathematics and Technical Non-mathematics scales reveals that the items dealing with scientific and technical subjects or activities are heavily weighted. Consequently some agreement between the two instruments would be expected. The extent of this agreement, however, is far from perfect. For example, there is not enough agreement to permit the prediction of a score on the Zyve Scientific Aptitude Test from a rating on the Technical Mathematics scale of the Strong. It is, therefore, concluded that the use of both of these instruments in counseling is better procedure than the use of either one or the other if the capacity of a student to do work in engineering or science is under consideration.

Received May 16, 1944.

References

1. Benton, A. L., and Perry, J. D. A study of the predictive value of the Stanford Scientific Aptitude Test (Zyve). *J. of Psychol.*, 1940, 10, 309-312.
2. Crawford, A. B. Review of the Stanford Scientific Aptitude Test in the *1940 Mental Measurement Yearbook*, 453–455. (Edited by O. K. Buros.)
3. Lindquist, E. F. *Statistical analysis in educational research.* N. Y.: Houghton Mifflin, 1940. Pp. 266.
4. Marshall, M. V. A study of the Stanford Scientific Aptitude Test. *Occupations*, 1942, 20, 433–434.
5. Seder, M. Group scales versus occupational scales for the Strong Vocational Interest Blank. In 1940 Achievement Testing Program in Independent Schools and Supplementary Studies, Educational Records Bull., No. 30, 51–56. N. Y.: Educational Research Bureau, 1940.
6. Strong, E. K., Jr. *Vocational interests of men and women.* Cal.: Stanford University Press, 1943. xxix, 1–746.
7. Thurstone, L. L., and Thurstone, T. G. *Manual of instructions, Psychological examination for college freshmen.* Washington, D. C.: American Council on Education. Published annually.
8. Zyve, D. L. A test of scientific aptitude. *J. educ. Psychol.*, 1927, 18, 525–540; see also the manual for the Stanford Scientific Aptitude Test published by Stanford Univ. Press, 1930.

The Measured Interests of Marine Corps Women Reservists *

Milton E. Hahn, Captain, USMCR, ** and Cornelia T. Williams, Major, USMCWR

Headquarters, U. S. Marine Corps, Washington, D. C.

A fundamental purpose of the whole Marine Corps' system of personnel classification is the assignment of each individual to the type of military duty where he or she can soonest and most efficiently serve the Marine Corps. In the accomplishment of this purpose, and as a preliminary to military assignment, each recruit is tested, interviewed, and given an opportunity to express his "choice of duty"; and all of the information thus obtained is recorded on his Qualification Card, Form 940. In this way, his basic aptitudes, his education and work experience, his special skills, and at least a crude indication of his interests become available as a basis for determining the most appropriate military assignment for each Marine.

Civilian occupational backgrounds of women enlisting in the Marine Corps supplied extremely useful data for proper assignment to military duty, but were at times misleading, or inadequate. For example, many school teachers and clerical workers joined the Marine Corps *to escape* from their civilian jobs. Many of the younger women joining the Corps had little, if any, previous work experience, and therefore, possessed no specific occupational skills. Others had considerable experience and had even developed a high degree of skill, but in a specialty so unrelated to any of the jobs open to women in the Marine Corps, that this experience was of little help in determining appropriate military assignments. Finally, many of the occupational specialties performed by women Reservists in the Marine Corps could not be significantly differentiated by available estimates or measures of basic aptitudes.

These facts made it obvious that in many instances a crucial factor in determining a military assignment would have to be a woman's *interest* in doing a particular type of work. This was especially true for women who had no special occupational skills, and for women who possessed occupational skills which could not be used directly in military duties. The need for a valid estimate of interests was equally great in selecting women for assignment to certain military jobs in which large numbers of women were needed but for which special training had to be arranged because so few women entering the service had had relevant previous experience.

---

* This report is part of a study made by the authors for the Classification Division, Detail Branch, Personnel Department, USMC, Headquarters, Washington, D. C. The opinions or assertions contained in this article are the private ones of the authors and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

** Dr. M. E. Hahn is now Director of the Psychological Services Center at Syracuse University, Syracuse, N. Y.

A search was necessary, therefore, for tools and techniques which could be utilized to bring about the assignment of women in the Marine Corps to duties for which they would not only possess the necessary minimum aptitudes and skills but in which they would also be interested and satisfied, and therefore, able to perform with greater efficiency.

In the early stages of the Marine Corps Women's Reserve program the only technique used for determining the interests of individuals was the interview. For most of the enlisted women this amounted to little more than one or two direct questions about "choice of duty" inserted at the end of the classification interview. A minority of each recruit class (those being considered for a few special types of military assignment) were reinterviewed to obtain a more precise description of their previous experience or to permit a clearer expression of their interests. Although the information relative to interests derived from interviews was helpful in making assignments to duty in the Marine Corps, a year of experience demonstrated the need for supplementary screening devices. Interviews, if they were to yield significant information on interests, were time-consuming; interviewers were variable in the reliability and validity of their judgments; and the claimed interests of the women, expressed merely as a preference for a specific military assignment, often had little or no validity as an indication of real interest in actual work activities.

A study of the situation was authorized by the Director of Personnel, Headquarters, United States Marine Corps, in March, 1944. The study included three major aspects: (1) the job satisfaction of female personnel in selected military occupations; (2) the measured interests of women in military occupations; and (3) a comparison of claimed and measured interests of women performing military duties in the Marine Corps. This report is concerned with the second aspect of the study, the measured interests of women now performing certain military duties.

It was decided to select for study a sample of women Reservists currently assigned to several widely different types of military duty. Matters of expediency limited the selection to those most readily available for testing. All of the women included in the study here reported were on duty either at Marine Corps Headquarters, Washington, D. C., or at the Marine Corps Air Station, Cherry Point, North Carolina.

Descriptive data on 667 enlisted women Reservists in the study are

Table 1

Descriptive Data Relative to 677 Enlisted Women, USMCWR, Who Accomplished the Kuder Preference Record

		Age in Years		Years of Education		Army Gen. Class Test—St. Score *		Army Mech. Apt. Test—St. Score *	
Military Duty	No.	Mean	Stand. Dev.	Mean	Stand. Dev.	Mean	Stand. Dev.	Mean	Stand. Dev.
Total	677	24.57	3.84	12.48	1.51	113.99	12.96	99.98	12.74
Aviation Machinist Mates	14	22.64	1.49	11.86	.99	112.71	9.30	104.93	14.60
Avn. Assembly and Repair (Mech.)	69	24.01	3.49	12.15	1.37	110.23	11.28	99.23	12.59
Motor Vehicle Operators	20	27.40	4.59	12.15	1.71	112.10	11.10	99.45	8.41
Cooks and Bakers	27	23.63	3.55	11.96	1.64	102.48	23.83	97.78	11.38
Syn. Train. Dev. (Instructors)	48	26.19	4.30	13.71	2.00	121.02	12.45	106.23	12.89
Stenographers	65	24.80	4.04	12.25	.68	117.77	11.57	97.39	12.16
Clerk-typists	161	24.22	3.47	12.22	1.25	112.30	11.62	97.96	12.04
Clerks, General	132	24.39	3.82	12.22	1.21	111.76	10.56	95.54	10.69
Duty NCO's	24	25.92	4.98	12.92	1.53	113.96	12.40	101.08	14.04
Others	117	24.49	3.58	13.12	1.78	119.21	12.19	106.97	12.55

* Norms for men are used for WR standard scale scores in the Marine Corps. GCT Forms C and D; MAT-3.

presented in Table 1.[1] (Three groups of women officers were also included in the study of measured interests but are not included in this table.) No marked differences are noted among groups in age or years of education. The mean age of motor vehicle operators is somewhat higher than the average age for the total group—27.4 vs. 24.6 years. The mean age for aviation machinist mates is somewhat lower than the mean age for the total enlisted group—22.6 vs. 24.6 years. Otherwise, the groups are relatively homogeneous in so far as age and years of education are concerned.

The Army General Classification Test, Forms c or d, and the Army Mechanical Aptitude Test, Form 3, are administered to all women Reservists. Norms for Army men are used. The minimum education requirement for enlistment in the Marine Corps Women's Reserve, completion of the twelfth grade or the equivalent, has resulted in a mean standard scale score of approximately 111 for enlisted women in the GCT. The mean GCT score for the enlisted sample discussed here is 114.0, somewhat above the mean of 102 obtained by enlisted men.

On the GCT, synthetic training devices instructors, stenographers, and the composite group of miscellaneous occupations labelled "Other" were above the group mean with respective means of 121.0, 117.8, and 119.2; standard deviations for these groups were similar to the one for the total group.

The only sub-group with a mean appreciably lower than that for the total group was Cooks and Bakers (mean 102.5). This group was also the most variable in the GCT scores.

Scores on the Army Mechanical Aptitude Test Form 3 showed no significant differences between means for the military occupational subgroups and the mean for the total sample except for aviation machinist mates. The $\frac{D}{\sigma_D}$ for the total mean and that for the aviation machinist mate sub-group was 3.93.

Civilian occupational backgrounds of these enlisted women were heterogeneous. The 161 women Reservists assigned to duty as clerk-typists, for example, came from 25 different civilian occupations.

Two interest tests or inventories were considered. The *Strong Vocational Interest Blank*, although it is the best validated and most widely used interest inventory, was impracticable from the standpoints of scoring costs and complexity of interpretation. The *Kuder Preference Record*,[2] although it is a comparatively new instrument, offered advantages in the simplicity of its administration, scoring, and interpretation. The *Kuder Preference Record,* therefore, was used to measure interests.

---

[1] Discrepancies in the number of cases reported between Table 1 and Table 2 are caused by certain records being unavailable for various reasons.

[2] The *Kuder Preference Record* Form BB, Science Research Associates, Chicago, Illinois.

The *Kuder Preference Record* contains nine scales which are purported to measure intensity of interest in nine broad areas of occupational activity. These nine areas are: (1) mechanical, (2) computational, (3) scientific, (4) persuasive, (5) artistic, (6) literary, (7) musical, (8) social service, and (9) clerical.[3]

This test is taken by punching with a stylus two of six possible choices for each item. Scoring is extremely simple. Single unit weights are used for scoring the items. The nine scores on the Preference Record yield an interest pattern or profile for the respondent. Adequate interest profile norms are not available for civilian occupational groups.

The test was standardized with the assumption that significant sex differences did not exist. The validity of this assumption can be tested by a comparison between the original norms, based upon a mixed group of college men and women, and the average scores of the Marine Corps Women Reservists. These comparisons are presented in Table 2. Although data on Kuder's original standardization group were not available for proper tests of the significance of differences between means, tentative

Table 2

Means and Standard Deviations for 791 USMCWR on the Nine Scales
of the Kuder Preference Record, Form BB ($N = 515$)

	Means		Standard Deviations		
Kuder Scales	USMCWR	Kuder Norms *	USMCWR	Kuder Norms *	$\bar{x} - \frac{D_\sigma}{\sigma_D}$
Mechanical	59.1	53.	20.18	18.	1.070
Computational	29.9	30.	12.62	10.5	.854
Scientific	57.8	51.	15.76	13.8	.828
Persuasive	61.	71.5	16.08	13.8	.835
Artistic	55.2	49.	15.74	16.3	.909
Literary	54.1	57.	15.39	15.5	.875
Musical	23.2	25.5	9.69	10.	.550
Social Service	78.1	72.	18.90	17.	1.006
Clerical	51.6	60.	16.48	15.5	.901

* Means and standard deviations for the Kuder Norm group were estimated from percentile norms on the Profile sheet based upon the assumption that distributions were normal.

---

[3] Manual for the *Kuder Preference Record,* Science Research Associates, Chicago, Illinois. Unfortunately, the present manual (June 1944) for the Preference Record is inaccurate and misleading. Those interested in the inventory are referred to the bibliography at the end of this report.

tests were made based on the assumption that distributions for Kuder's norm group were normal. These data are presented in Table 3. The null hypothesis was not refuted.

Table 3

Decile Norms for 791 Adult Women, USMCWR, on the Nine Scales of the Kuder Preference Record

	Raw Score										
	Deciles										
Kuder Scales	10	20	30	40	50	60	70	80	90	100	Range
Mechanical	33	40	46	52	57	63	71	78	86	109	21–115
Computational	14	18	22	26	29	32	35	39	47	65	5– 69
Scientific	36	43	48	53	57	61	65	71	79	96	26–102
Persuasive	41	47	51	55	59	64	69	74	81	108	29–112
Artistic	34	41	46	50	54	58	63	69	76	95	21– 98
Literary	34	40	44	49	53	57	62	67	75	93	23– 96
Musical	10	14	17	20	23	26	29	32	36	46	3– 71
Social Service	53	60	67	73	78	83	89	95	102	120	30–122
Clerical	31	36	41	45	51	56	60	66	73	97	17–100

Table 3 contains decile raw score norms for the 791 women who completed the Preference Record. Intercorrelations were computed for the nine scales included in the inventory. These data are presented in Table 4. With five exceptions—mechanical vs. scientific, mechanical vs. literary, mechanical vs. clerical, computational vs. clerical, and artistic vs. social service—there is little evidence that correlations among the scales depart markedly from zero in so far as this sample is concerned.

Military occupational sub-samples of women Reservists in the Marine Corps were compared statistically. These sub-samples were: Officer, line; Officer, staff; Officer, technical specialty; Aviation Machinist Mate; Assembly and Repair (Aviation); Motor Vehicle Operator; Cooks and Bakers; Synthetic Training Device Instructors; Stenographer; Clerk-typist; Clerk, general; Duty Non-Commissioned Officer; and Other.[4]

Means and standard deviations for these military occupational groups are presented in Tables 5 and 6 respectively.

Time did not permit a comparison of the mean of each occupational group with the mean for every other occupational group. The mean for each occupational group was, however, compared with the mean for the total sample. The Preference Record did differentiate most of the occu-

---

[4] The "Other" group consisted of 122 cases in military occupations the samples for which were too small or the occupation of too little importance to warrant separate treatment. Distributions for this occupational composite closely approximated those for the total sample.

Table 4

Intercorrelations among the Nine Scales of the Kuder Preference Record for a Sample of 791 Adult Women (MCWR)

	Mech.	Comp.	Sci.	Pers.	Art.	Lit.	Mus.	Soc. Ser.	Cler.
Mechanical		−.068	.404	−.270	.158	−.349	−.299	−.160	−.360
Computational	−.068		.240	−.287	−.300	−.091	−.161	−.146	.483
Scientific	.404	.240		−.403	−.120	−.274	−.323	−.022	−.253
Persuasive	−.270	−.287	−.403		−.166	.220	.030	.042	.017
Artistic	.158	−.300	−.120	−.166		−.110	.093	−.349	.291
Literary	−.349	−.091	−.274	−.220	−.110		.109	−.231	−.008
Musical	−.299	−.161	−.323	.030	.093	.109		−.197	−.161
Social Service	−.160	−.146	−.022	.042	−.349	−.231	−.197		−.206
Clerical	−.360	.483	−.253	.017	−.271	−.008	−.166	−.206	

An r of .112 is significant at the 1% level.

An r of .085 is significant at the 5% level.

Table 5

Mean Scores for 12 Occupational Groups of Women, MCWR, on the Nine Scales of the Kuder Preference Record

		Mean Scores								
		Kuder Scales								
Occupational Group	No.	Mech.	Comp.	Sci.	Pers.	Art.	Lit.	Mus.	Soc. Ser.	Cler.
Total WR Sample	791	59.12	29.85	57.57	60.95	55.20	54.11	23.18	77.97	51.61
Officer, Line	23	60.22	28.17	56.39	60.78	51.74	57.61	20.61	88.57	45.22
Officer, Staff	20	55.05	31.70	59.80	61.80	57.55	57.65	22.65	76.25	46.65
Officer. Tech. Spec.	17	69.59	30.12	63.00	57.94	65.18	52.65	18.35	71.00	44.18
Aviation Machinist Mate	14	71.14	23.36	69.93	57.36	52.43	55.86	20.57	76.57	43.50
Assembly and Repair	75	71.45	28.28	61.31	55.55	59.60	46.47	20.12	78.29	46.85
Motor Vehicle Operator	20	59.15	30.55	59.55	58.10	49.05	46.30	24.10	87.95	54.05
Cooks and Bakers	29	61.14	29.52	55.97	62.24	49.83	55.17	21.97	82.69	50.76
Syn. Tr. Dev. Opr.	60	60.00	30.80	58.88	61.68	57.98	57.60	23.82	74.23	48.10
Stenographer	76	52.86	29.32	55.96	63.40	51.79	55.82	23.58	75.07	59.76
Clerk-typist	167	56.35	30.59	57.50	59.33	53.96	55.98	24.23	76.92	54.43
Clerk, General	143	56.90	30.00	54.41	63.27	55.72	53.17	24.52	78.04	54.22
Duty NCO	25	53.32	27.32	52.56	71.84	53.40	51.52	21.06	85.92	52.32
Total	669									

Table 6

Standard Deviations for 12 Occupational Groups of Women, MCWR, on the Nine Scales of the Kuder Preference Record

		Standard Deviations								
		Kuder Scales								
Occupational Group	No.	Mech.	Comp.	Sci.	Pers.	Art.	Lit.	Mus.	Soc. Ser.	Cler.
Total WR Sample	791	20.18	12.62	15.76	16.08	15.74	15.39	9.69	18.90	16.48
Officer, Line	23	24.94	13.28	17.85	19.22	17.12	14.56	8.47	16.79	14.66
Officer, Staff	20	15.99	13.79	13.17	22.58	12.28	14.71	8.55	18.25	12.91
Officer, Tech. Spec.	17	23.50	12.08	15.17	16.88	15.33	17.03	10.01	17.09	14.55
Aviation Machinist Mate	14	17.09	8.82	12.34	10.34	12.70	15.85	7.73	19.26	17.48
Assembly and Repair	75	18.74	10.54	15.31	14.81	16.74	11.69	8.75	17.42	16.89
Motor Vehicle Operator	20	17.56	12.29	11.47	15.32	15.70	12.82	9.35	12.68	16.34
Cooks and Bakers	29	18.25	10.61	17.25	13.52	14.95	14.10	9.94	17.94	14.67
Syn. Tr. Dev. Opr.	60	19.88	12.76	16.01	16.94	13.96	15.08	10.46	16.29	15.09
Stenographer	76	16.76	11.47	17.05	15.02	15.35	16.78	9.18	19.50	19.07
Clerk-Typist	167	20.56	13.15	14.91	14.38	15.20	15.80	9.56	19.58	16.06
Clerk, General	143	19.31	14.10	15.87	15.58	15.91	14.93	9.71	19.59	16.47
Duty NCO	25	22.51	11.94	10.12	17.46	16.85	14.77	7.97	21.31	13.66
Total	669									

Table 7

Significance of the Differences between Means ("*t*") of Each of 13 Occupational Sub-Samples and a Sample of 791 Adult Women (MCWR) on the Nine Scales of the Kuder *Preference Record*

Occupational Sub-Group	*N*	Mech.	Comp.	Sci.	Pers.	Art.	Lit.	Mus.	Soc. Ser.	Cler.
Officer, Line	23	(−) .25	.63	.35	.05	1.04	(−)1.08	1.26	(−)*2.66*	1.84
Officer, Staff	20	.89	(−) .64	(−) .63	(−) .23	(−) .66	(−)1.02	.24	.40	1.34
Officer, Tech. Spec.	17	(−)*2.11*	(−) .09	(−)*1.41*	.76	(−)*2.58*	.39	*2.05*	1.51	1.84
Avn. Mach. Mate	14	(−)*2.21*	1.92	(−)*2.92*	.83	.66	(−) .42	1.00	.27	1.82
Avn. Assembly and Repair	75	(−)*5.08*	1.04	(−)*1.97*	*2.80*	(−)*2.30*	*4.18*	*2.63*	(−) .14	*2.38*
Motor Vehicle Operator	20	(−) .01	(−) .24	(−) .56	.78	1.73	*2.25*	(−) .42	(−)2.35	(−) .65
Cooks and Bakers	29	(−) .53	.14	.53	(−) .43	1.81	(−) .37	.66	(−)1.32	.28
Syn. Tr. Dev. Instr.	60	(−) .32	(−) .56	(−) .62	(−) .34	(−)1.33	(−)1.69	(−) .49	1.49	1.60
Stenographer	76	*2.62*	.36	.84	(−)1.27	1.81	(−) .91	(−) .35	1.27	(−)*4.06*
Clerk-Typist	167	1.61	(−) .68	.05	1.20	.93	(−)1.42	(−)1.28	.65	(−)*2.01*
Clerk, General	143	1.22	(−) .13	*2.20*	(−)1.60	(−) .36	.68	(−)1.52	(−) .04	(−)1.74
Duty NCO	25	1.41	.99	*4.98*	(−)*3.32*	.56	.83	1.07	(−)*2.06*	(−) .21
Miscellaneous, Occ.	122	(−) .32	(−) .71	(−) .39	.27	(−) .66	(−) .53	(−) .43	.10	*2.35*

Note: d.f. for all "*t*"s is > 30. A "*t*" of 2.58 is significant at the 1% level. A "*t*" of 1.96 is significant at the 5% level. The negative sign in each instance signifies that the mean score for the occupational sub-sample was *above* the average score for the total group.

Table 8

Means and Standard Deviations for Three Groups of Satisfied *vs* Dissatisfied Clerical Workers MCWR

Interest Scales, Kuder Preference Record	Stenographers				Clerk-Typists				Clerks, General			
	Means		SD's		Means		SD's		Means		SD's	
	High *	Low †	High	Low	High	Low	High	Low	High	Low	High	Low
Mechanical	52.2	52.1	16.9	19.9	48.1	61.1	16.1	23.1	56.3	57.9	19.2	19.6
Computational	34.4	26.4	11.7	12.7	34.0	28.5	12.4	16.3	36.9	27.8	16.0	13.7
Scientific	53.6	62.5	17.2	17.6	57.1	54.9	12.5	15.5	57.0	53.9	18.2	15.3
Persuasive	64.2	62.4	19.2	9.2	57.6	63.5	16.7	14.9	61.5	64.0	13.9	16.6
Artistic	52.2	52.2	18.0	13.7	50.0	53.0	12.9	15.3	55.1	56.5	15.3	18.1
Literary	60.0	54.7	15.7	14.7	56.0	54.7	14.7	15.0	51.4	55.2	13.6	16.2
Musical	20.6	25.6	7.7	8.0	24.4	25.9	11.1	8.9	21.7	25.0	9.8	9.3
Social Service	69.1	57.8	21.1	17.2	78.6	79.5	20.0	18.7	78.9	77.1	22.2	18.7
Clerical	66.1	47.9	18.7	16.6	65.6	55.5	13.1	17.8	60.1	51.8	15.8	16.2
Number	25	14			23	50			27	72		

* High = High Job Satisfaction.

† Low = Low Job Satisfaction.

Table 9

Significance of Differences ("*t*") between the Means of Satisfied and Dissatisfied Clerical Workers, MCWR

Occupational Sub-Group	*N*	Kuder Keys								
		Mech.	Comp.	Sci.	Pers.	Art.	Lit.	Mus.	Soc. Ser.	Cler.
Clerk Typists High *vs.*	23	(−)*2.40*	1.43	.59	(−)1.49	(−).79	.34	(−) .64	(−) .18	*2.41*
Low Job Satisfaction	50									
Stenographers High *vs.*	25	.01	1.91	(−)1.51	.32	(−).01	1.01	(−)1.88	(−)1.55	*2.97*
Low Job Satisfaction	14									
Clerks, General High *vs.*	27	(−) .36	*2.76*	.84	(−) .71	(−).36	(−)1.08	(−)1.53	.39	*2.28*
Low Job Satisfaction	72									

Note: d.f. for all "*t*"s is > 30. A "*t*" of 2.58 is significant at the 1% level. A "*t*" of 1 96 is significant at the 5% level. "*t*"s preceded by (−) indicate that the mean for the "dissatisfied" group is larger.

pational groups from generality on one or more scales despite the fact that data for the sub-group were not removed from the generality for the comparisons. Occupational groups not differentiated significantly from generality by at least one of the nine scales were officer—staff, cooks and bakers, and synthetic training devices instructors. Table 7 presents "*t*'s" for differences between means for each occupational sub-group and the total sample for each of the nine interest areas.

Table 7 is of particular interest because it illustrates a phenomenon which has not received much attention in the literature concerned with the measurement of interest—"rejection" or "aversion" scores. The topic is not particularly germane to this report, but attention is called to the fact that occupational groups are as clearly differentiated on interest scales by their "rejection" scores as they are by their "acceptance" or "interest in" scores. Bingham's statement that a C grade obtained on an occupational key of the Strong Vocational Interest Blank means "No" may need change to "No, for that occupation; yes, for some others."[5]

The *rejection* of items positively weighted for the clerical scale may be as important to the measurement of an interest in the duties of an Aviation assembly and repair worker as the *acceptance* of items positively weighted for the mechanical scale.

The preliminary aspect of this investigation was a study of job satisfaction which indicated satisfied and dissatisfied groups of workers. Scores on the Preference Record scales were separated for these satisfied and dissatisfied groups. Scores on the inventory for these two groups showed marked differences between them on one or more interest scales. Tables 8 and 9 present these data for three groups of clerical workers. These tables make it obvious that occupational norm groups for a test of interests should be composed only of those who are satisfied with their jobs. Groups so constituted in this study have interest profiles much more clearly differentiated from the total standardization groups than sub-samples containing a large proportion of individuals disinterested in or dissatisfied with the type of work they are doing.

Conclusions

Although the major purpose of this report is to present norms, and statistics related to these norms, certain conclusions of general interest are presented here.

---

[5] Bingham, W. V., *Aptitudes and aptitude testing*, Harpers, 1937. Appendix, Section IX, p. 356.

[6] None of the other military occupational sub-samples contained enough dissatisfied workers to make an analysis of this sort possible or necessary.

1. Certain occupational sub-samples of women Reservists in the United States Marine Corps can be differentiated by scores on various scales of the *Kuder Preference Record.*

2. In the case of three sub-groups of clerical workers—stenographers, clerk-typists, and clerks, general—certain scales on the Preference Record differentiate satisfied from dissatisfied workers. In each of these three sub-groups, the satisfied workers are more clearly differentiated from the total group of 791 than are sub-groups containing both satisfied and dissatisfied clerical workers.

3. "Rejection" (low or negative) scores on certain scales differentiate occupational groups from the generality as markedly as do the "acceptance" (high or positive) scores.

4. Group differentiation is a matter of patterns or profiles which are characterized by both "acceptance" and "rejection" scores.

5. Comparison of individual profiles with occupational sub-sample profiles permits a surprisingly good interest screen for use in the assignment of women Reservists to military duty.

Received July 5, 1944.

References

1. Crawford, A. B. Review of the preference record in *The 1940 Mental Measurement Yearbook*, Edited by Oscar K. Buros. Highland Park, New Jersey, 1941, pp. 447–449.
2. Kuder, G. Frederic. The stability of preference items. *J. Soc. Psychol.*, 1939, 19, 41–50.
3. Traxler, A. E. A note on the reliability of the revised Kuder Preference Record. *J. appl. Psychol.*, 1943, **27**, 510–511.
4. Traxler, Arthur E., and McCall, William C. Some data on the Kuder Preference Record. *Educ. & Psychol. Meas.*, 1941, 1, 253–268.
5. Wittenborn, J. R., Triggs, Frances O., and Feder, Daniel D. Comparison of interest measurement by the Kuder Preference Record and the Strong Vocational Interest Blanks for men and women. *Educ. & Psychol. Meas.*, 1943, 3.

Personality Patterns of Adolescent Girls: I. Girls Who Show Improvement in IQ *

Dora F. Capwell

Trainee Acceptance Center, Public Schools, Pittsburgh, Pa.

The research described in this report was designed to discover what differences there are between those adolescent girls whose IQ's change significantly upon retest and those whose IQ's show only slight changes. Its purpose, therefore, is to throw additional light upon the problems of individual, clinical diagnosis, which inevitably involves or at least implies prediction of a future level of performance. The determination of what factors within the individual affect test scores is a more narrowly defined problem than the one of mere constancy of IQ, measured in terms of group data, and is, of course, significantly related to accuracy of diagnosis, which is preliminary to the planning of effective educational and social treatment.

Literally hundreds of studies have been reported which contain retest data from intelligence tests. They have been reviewed and summarized by Baldwin (2, 3), Burks (6), Foran (9, 10), Nemzek (24), and Thorndike (31). Although the studies have dealt with all age levels, various time intervals, several different tests, and a wide variety of testing conditions, the coefficients of correlation between test and retest have ranged from .63 to .95, most groups showing correlations of .84 or better. Treated by the correlation technique, the data consistently showed high positive correlation between test and retest, and hence the authors assumed that they demonstrate the relative constancy of the IQ.

The studies which are of special interest here are the ones which attempt to determine the causes of change in level of test performance in those cases which showed variability of IQ. These may be divided into three groups: studies which include special analysis of the extreme cases in the distribution of IQ changes (7, 13, 20, 22, 25, 32), studies made on selected groups of problem children (5, 12), and studies of the effect of specific factors, such as attitude, emotion, and psychopathic personality (5, 11, 16, 19, 23, 27, 28). All of these studies have been characterized by a lack of objective data aside from the intelligence test scores. Most frequently causes of IQ variability have been assigned by surveying the history, considering the events which have taken place between tests, and making a highly subjective judgment. No study reported convincing evidence that personality factors are functionally associated with changes in IQ on retest, although many authors drew that conclusion from what is here considered insufficient evidence.

---

* The writer is indebted to the Department of Psychology, University of Pennsylvania, and particularly Drs. Malcolm G. Preston, Francis W. Irwin, and Miles Murphy, who served as research advisers in the final stages of the work, for advice and constructive criticism.

Gratitude is due to the Bureau of Psychological Services, State of Minnesota, which sponsored the study. Valuable assistance was given by the late Dr. Fred Kuhlmann and his successor, Dr. Stuart Cook. The staff of the State School for Girls at Sauk Centre, Minnesota, and the Sauk Centre Public Schools gave their wholehearted cooperation.

Dr. Starke R. Hathaway, University of Minnesota, made the Multiphasic Personality Inventory available to the writer before its publication and was helpful throughout the period of collecting data.

The present study attempts to present a more objective description of the individuals studied by utilizing scores on a variety of tests, analyzing items of possible significance in the history, and also analyzing the record of personal-social events which occurred between examinations. The use of both a normal group and a definitely maladjusted group permits a more thorough exploration of the significance of a wide variety of concomitant factors. Although the study was planned to investigate factors related to IQ *variability*, the results led to a study of those whose IQ *improved* as compared with those whose IQ remained relatively constant. Three major questions will be asked of the data.

(1) Are there significant differences of personality and experiential background between those adolescent girls whose IQ shows significant improvement and those whose IQ changes only slightly?

(2) Is significant improvement in the IQ of adolescent girls related to improvement in the adjustment pattern of the total personality?

(3) Are the relationships which are investigated first in a delinquent, institutionalized, adolescent, female group demonstrable in a non-delinquent, adolescent, female group in the community?

Procedure

Subjects. The subjects were 101 delinquent girls and 85 non-delinquent girls who were between ages 12 and 18 and whose IQ on the first test was not less than 60. The delinquents were consecutive admissions to the Minnesota State School for Girls, beginning in September, 1941. The non-delinquents were in the consolidated public school at Sauk Centre, Minnesota, and were chosen from grades which would match the usual grade distribution of girls entering the State School. The principal selected every other girl from the grade lists until the necessary number was obtained. The groups were roughly equated for urban-rural backgrounds. Both schools have a population which is about two-thirds urban and one-third rural, when their places of residence are classified according to the U. S. Census criterion of urban and rural.

The method of selection resulted in two groups of the following description. The delinquents ranged in age from 13 to 19 with a median age of 16. The non-delinquents ranged from 12 to 18 with a median age of 15. The slight difference in age range is a reflection of the fact that the delinquents show more school retardation, so that their average age for each grade is a little older. The median grade placement of the delinquents was 9th grade; 86% were in grades 7 to 10, inclusive, 12% were above 10th grade, and 2% were below 7th grade. The median grade placement of the non-delinquents was also 9th grade, with 94% in grades 7 to 10, inclusive, and 6% above 10th grade.

Collection of Data. The data consist of test scores obtained on two psychological examinations, items of classification from the personal-social history, and items of classification related to progress between examinations. The delinquents were examined the first time within their initial two weeks at the School, and the second examination was given from 6 to 15 months later. The non-delinquents were examined the first time in the fall of 1941 and were re-examined from 4 to 13 months later. The tests given at both examinations were the Kuhlmann Tests of Mental Development (18), the Minnesota Multiphasic Personality Inventory (14), the Washburne Social Adjustment Inventory (33), and the Pressey Interest-Attitude Test (26). Three tests which each subject took only once were the Terman-Miles Test of Masculinity-Femininity (30), the Vineland Social Maturity Scale (8), and the Stanford Achievement Test (17). Three testing sessions were used to complete the entire battery for each examination. All of the tests were administered by the writer with the exception of the delinquents' Stanford-Achievement Tests, which are given routinely by the school principal at the State School.

From the personal-social history of the delinquents tabulations were made of the occupation and education of parents, national and racial background, language spoken in home, work experience, length of time out of school, type of delinquency, and other social problems within the family. Information on the non-delinquents was obtained by personal interviews and from the school records regarding occupation and education of parents, language spoken in the home, work experience, and any social problems in the family. For the delinquents there were also Home Ratings made by the State School field workers on a large proportion of the cases.

For the interval between tests the record of the delinquent cases included health status, school grades, discipline reports, and ratings of work habits and general conduct made by the housemother, work supervisor, and supervisor of home life. For the non-delinquents there was a record of health history and school grades. The latter were recorded from the principal's records, and the student's report on health was checked by records in the office of the school nurse.

Results

Intelligence Test Results. The Kuhlmann Tests of Mental Development were used as the measure of intelligence, and, hence, the main characteristics of these tests should be kept in mind when examining the results. When applied at the age levels used in this study, the scale consists of sixteen tests, all of which are timed, and all of which are scored at successively higher levels, depending upon the amount accomplished in a given time interval, usually two minutes. Five of the tests demand the use of spatial relations and of spatial imagination. The others are more strictly verbal. Each test is given a time score and an accuracy score as well as a mental age score. Sixteen is the highest chronological age denominator used in computing IQ's. The results of these groups are summarized in Table 1.

Table 1

Kuhlmann Tests of Mental Development

	Delinquents			Non-Delinquents		
	Mean IQ	S.D.	σm	Mean IQ	S.D.	σm
1st Test	87.40	17.10	1.70	101.88	17.59	1.90
2nd Test	95.65	19.62	1.95	111.76	20.50	2.22

The non-delinquents are significantly brighter than the delinquents on both tests. $D/\sigma D$ *on the first test is 5.70, and on the second one it is 5.46.* Despite the difference in mental level, each group showed about the same amount of shift in IQ on the second test. The average amount of change was an *increase* of 8 points for the delinquent group and 10 points for the non-delinquent group.

The coefficient of reliability was computed with the test-retest scores for each group. Product moment correlations also were computed to determine the relationship of IQ changes to time interval, change in speed score, change in accuracy score, and level of first IQ (Table 2). The time interval between tests and the change in accuracy score had a negligible relationship to the changes in IQ, but there was a marked relationship between speed score and changes in IQ. The reliability coefficient is high enough to be considered satisfactory for a test used for individual diagnosis. It is of interest in passing that these are the first reliability coefficients reported on the Kuhlmann Tests of Mental Development.[1] The correlation between IQ change and level of first IQ is so low that no relationship between them is indicated, implying that the distribution of changes at various IQ levels is not affected by the IQ level itself.

Table 2

Product-Moment Correlations Based on the Results from the Kuhlmann Tests of Mental Development

	Delinquents	Non-Delinquents
Test-retest reliability	.90 ± .01	.88 ± .01
IQ change and time interval	−.02 ± .06	.06 ± .07
IQ change and change in accuracy score	.18 ± .06	.15 ± .07
IQ change and change in speed score	.44 ± .05	.41 ± .06
IQ change and level of first IQ	.08 ± .06	.15 ± .07

We find that the IQ's of approximately 36% or 36 of the 101 delinquents changed more than 10 points, and 48% or 41 of the 85 non-delinquents changed more than 10 points.[2] Inasmuch as the standard error of a score is 5.4 for the delinquents and 6.7 for the non-delinquents, a change of more than 10 points would occur by chance only 5 times in 100 for the delinquents and 10 times in 100 for the non-delinquents. A change of more than 10 points, then, with this group may be considered a significant change. In presenting the results of the other tests, we shall divide each major group into two subgroups—those whose IQ changed 10 points or less, called the constant group, and those whose IQ changed more than 10 points. These latter groups should be thought of as variable groups in contrast to the constant ones, but since only one delinquent and one non-delinquent had IQ's which regressed more than 10 points, the remainder really improved, and the groups are essentially groups which showed significant improvement in IQ as against those whose IQ did not improve. Hence, the variable groups will be designated the improved groups.

Achievement Test Results. In the case of all three achievement scores the delinquents have a lower grade level of achievement than the non-delinquents, and in *both* of the major groups the constant IQ group has a lower achievement level than the improved group. Table 3 demonstrates these relationships, not in terms of absolute achievement test scores, but in terms of the relation of the individual's achievement level to his actual grade level. The actual grade level of the non-delinquents was figured on

---

[1] For a discussion of why they were omitted in the original report of the Kuhlmann Tests, see (18), pp. 16–17.

[2] $D/\sigma Dp$ for those two percentages is 1.71.

Table 3

Difference between Actual Grade Level and Achievement Grade Score *

Score	Delinquents †					Non-Delinquents				
	Constant Group $N = 61$		Improved Group $N = 36$		$D/\sigma D$ between Constant and Improved Groups	Constant Group $N = 44$		Improved Group $N = 41$		$D/\sigma D$ between Constant and Improved Groups
	Mean Diff.	S.D.	Mean Diff.	S.D.		Mean Diff.	S.D.	Mean Diff.	S.D.	
Total Score	−1.41	1.39	−1.12	1.40	1.03	−1.14	1.55	−.47	1.35	2.16
Reading score	− .75	1.57	− .17	1.34	1.93	− .81	1.51	−.11	1.31	2.33
Arithmetic Score	−2.15	1.66	−1.64	1.87	1.37	−1.65	2.02	−.89	1.95	1.81

* 1.00 equals one total grade.

† Four delinquents were absent when the tests were given.

the basis of the time of the school year when the test was given, and in the delinquent group it was figured as closely as possible by the grade placement at the time of the year when the girl left school. The difference between actual grade level and the achievement test score gives a "difference achievement score." These have been averaged for each group and are presented in Table 3. When compared with the norms for the Stanford Achievement Test, each group shows some retardation in grade achievement. It should be recognized, however, that this is true partly because the ceiling of the test is too low for a few of the group who were in 10th, 11th, and 12th grades and who made scores above the maximum grade level of the test. No doubt the critical ratios between delinquents and non-delinquents would be slightly higher if the norm did not stop at the 11.0 grade.

Personality Test Results. The results of the other tests used in the battery provide data for answering the first question posed; namely, are there significant differences of personality, as measured by this group of personality tests, between those adolescent girls whose IQ's show significant improvement and those whose IQ's show only slight variability?

The Minnesota Multiphasic Personality Inventory is scored on eleven scales, three of which are checks on the validity with which the subject has answered the Inventory. These are the scores for "?'s," L, and F. The "?" score is the sum of items put in the "Cannot say" category rather than answered as true or false. L is a lie score, which if too high, shows that the subject is attempting to present too favorable a picture of himself. The F-score is a check on how many extremely unfavorable items are included in the score; it includes items which normally are affirmed by only a few cases in a thousand. Although the F-score tends to go up as maladjustment increases, an extremely high F-score suggests carelessness, lack of comprehension, deliberate falsification, or scoring errors. The other eight scales are measures of specific abnormal tendencies, grouped under familiar psychiatric classifications. Their letter abbreviations have the following meanings: Hs—hypochondriasis, D—depression, Hy—hysteria, Pd—psychopathic deviate (formerly called psychopathic personality), Pa—paranoia, Pt—psychasthenia, Sc—schizophrenia, and Ma—mania.

Table 4 shows the significance of difference between the subgroups of each major group. The ratios do not indicate a significant difference between the results for the constant and improved groups, but the test did discriminate extremely well between the delinquent and non-delinquent groups. The differentiation which this scale and the other personality tests made between delinquents and non-delinquents will be presented in a subsequent report.

Table 4

Minnesota Multiphasic Personality Inventory—Significance of Differences of Raw Scores

Scale	Delinquents, Diff. between Improved and Constant Groups		Non-Delinquents, Diff. between Improved and Constant Groups	
	First Test *D/σD*	Second Test *D/σD*	First Test *D/σD*	Second Test *D/σD*
?-Score	.12	.55	.27	1.03
L-Score	.48	.01	.71	1.00
F-Score	3.30	2.07	.76	1.24
Hs-Scores	.26	1.96	.35	.53
D-Scores	.96	.40	.61	1.94
Hy-Scores	1.75	.92	1.06	1.13
Pd-Scores	.30	.62	.22	.24
Pa-Scores	.61	1.60	.25	.67
Pt-Scores	.00	1.35	.43	.64
Sc-Scores	.43	1.18	.50	.14
Ma-Scores	1.20	.13	2.25	2.03

The same treatment has been given the scores from the other personality tests, namely, the Washburne Social Adjustment Inventory, the Pressey Interest-Attitude Test, and the Terman-Miles Test of Masculinity-Femininity. The Vineland Social Maturity Scale is a different type of test from these others, but may be grouped with them, and Table 5 shows the significance of differences between each group on both examinations. The Washburne, the Terman-Miles, and the Vineland do not show significant differences between the constant and improved groups. That the Pressey shows high critical ratios between the constant and improved delinquents but not the non-delinquents and does not discriminate be-

Table 5

Other Personality Tests—Significance of Difference of Raw Scores

Test	Test No.	Delinquents, Diff. between Constant and Improved Groups *D/σD*	Non-Delinquents, Diff. between Constant and Improved Groups *D/σD*
Washburne	1st	.60	1.06
Washburne	2nd	.40	.06
Pressey	1st	2.79	.15
Pressey	2nd	3.59	.12
Terman-Miles	(one only)	1.22	.78
Vineland *	(one only)	.49	.21

* *D/σD* between social quotients.

tween the total group of delinquents and non-delinquents appears to be a chance phenomenon. The Terman-Miles shows no significant differences between any groups.

Personal History. The second part of question (1) is: are there significant differences of background between those adolescent girls whose IQ showed significant improvement and those whose IQ varied only slightly? The first item on which they are to be compared is the occupational level of the father. In cases where the father is deceased or not with the family and the mother works, the occupation of the mother was rated. The occupations were classified according to the Taussig (29) scale, which has been used by Hildreth (15) and others. Chi square was computed, in a five by two table, to find out if there is a reliable difference in the distribution of occupational level of parents between delinquents and non-delinquents and the constant and improved subgroups of each. Between parental occupations of delinquents and non-delinquents P-value is less than .01. Application of chi square to constant and improved delinquent groups yields a P-value of .60, and applied to constant and improved non-delinquent groups the P-value is .04.

The indications are, therefore, that there is a reliable difference between the occupational level of the parents of the delinquents and non-delinquents, the occupational level being higher for the parents of non-delinquents. There is not a reliable difference between occupational levels of the constant and improved delinquent groups, but with the non-delinquents the difference approaches significance, the parents of the improved group tending to have a higher occupational level than those of the constant group.

The parents' education is another item which was investigated as far as possible with each group, but the educational information is more limited than the occupational information. In only 47 of the 101 delinquent cases and 77 of the non-delinquents was it possible to obtain information regarding parents' education. The chi square test was applied in a four by two table and yielded a P-value between .01 and .02 between the education of the delinquents' parents and the parents of the non-delinquents, indicating a reliable difference in favor of the non-delinquents. The subgroups within the two larger groups are so small, due to the incompleteness of the data, that chi square is not an appropriate test of differences, and for the same reason more elaborate treatment does not seem indicated.

Differences in race, nationality, and language were so small as to be insignificant in all groups. Only nine of the delinquents and one non-delinquent had one or both parents of foreign birth. Five of each major group were not of the white race. Thirteen delinquents and two non-delinquents came from bi-lingual homes. These few girls were about evenly distributed between the constant and improved groups. Work experience is another factor which differed markedly between delinquents and non-delinquents, the delinquents having worked more, but it had no apparent significance with respect to the two IQ groups.

By means of the records at hand and personal interviews an attempt was made to tabulate any social problems within these girls' families, including problems of health, mental retardation, mental instability, and social maladjustment, such as delinquency or criminality. Of the delinquents 72% of the constant IQ group had social problems within the family and 72% of the improved group. Of the non-delinquents 23% of the constant group and 20% of the improved group had family problems of the types mentioned. Again the same pattern is seen wherein there is conspicuous difference between delinquents and non-delinquents but little or none between subgroups of each.

Three additional items were tabulated for the delinquents only. The type of delinquency was not related to IQ changes, inasmuch as 91% of the constant group and 94% of the improved group were committed for sex delinquency. Ratings on the home conditions of the delinquents were made by field workers from the State School who visit the home of each girl who has been committed. They used an adaptation of the Whittier Scale for Grading Home Conditions (34), which includes ratings on Necessities, Neatness, Size, Parental Conditions, and Parental Supervision. The ratings were examined to see if those girls whose IQ improved after a period in the institution came from the poorest or most unsatisfactory homes. There were no striking differences between the homes of the girls in the two subgroups. Ratings on the first three categories covered the entire range, but on the latter two, Parental Conditions and Parental Supervision, there was piling up on the unfavorable end of the scale. The third item considered for the delinquents only was the length of time out of school prior to commitment. The Kuhlmann Tests are pencil-and-paper tests, much like school work, and it was thought that a girl who had been out of school for some time might find it easy to better her score after a return to school. However, classifying the girls' length of time out of school in a four part table and applying the chi square test, a P-value of between .70 and .80 was obtained, so there is no evidence that length of time out of school prior to taking the tests had any relation to ability to improve the score.

Personality Tests (Changes from First to Second Examination). The second major question to be answered is whether significant improvement in the IQ of adolescent girls is related to improvement in the adjustment pattern of the total personality. Our data for answering this question are the amount of change in the scores of those three personality tests which were given twice and the relation of these changes to changes in IQ. A second type of data is the analysis of personal-social events between examinations.

The amount of change in scores on the personality tests was averaged for the various groups and compared. Table 6 shows that changes in personality adjustment as measured by scores on this group of personality tests occurred no more frequently in the improved groups than the constant groups. More detailed examination of the score changes shows that there was a slight but consistent tendency for the delinquents' scores to shift more toward the mean for the normal population on the second test, but the average change was so slight and also so evenly distributed with relation to constant and improved groups that it is not of significance for the present problem.

Table 6

Significance of Difference between Mean Changes of Score on 1st and 2nd Personality Tests

Test	$D/\sigma D$ Changes of Constant and Improved Delinquents	$D/\sigma D$ Changes of Constant and Improved Non-Delinquents
Minnesota Multiphasic		
?-Score	1.02	.53
L	.40	.33
F	1.30	.58
Hs	1.90	.21
D	1.63	2.05
Hy	1.17	2.00
Pd	1.28	.20
Pa	1.82	.31
Pt	1.49	.39
Sc	.83	.10
Ma	1.80	.35
Washburne S.A. Inventory	1.38	2.40
Pressey Interest-Attitude	.87	.09

Personal-Social History in Interval between Tests. Health frequently is mentioned as a cause of changes in test performance. The health record for the delinquents was available at the School. With the non-delinquents a classification was made on the basis of the report of each girl on number of illnesses, which was checked by records of the school nurse, who attempts to record causes for each absence resulting from illness. In the delinquent groups 23% of the constant IQ group and 45% of the improved IQ group received treatment for major health conditions, including pregnancies. Chi square applied to all classifications yielded a P-value between .10 and .05, and $D/\sigma D$ between the major treatment categories of each group was 2.24. Although there is little likelihood of of the difference occurring by chance, neither test quite meets the criterion for a reliable difference. Twenty-six cases in the entire group were fitted with glasses between examinations. They were about evenly divided between the constant and improved groups, but their average IQ change, however, was 11 points, which is 3 points higher than the average.

The chi square test applied to the classification of health history for the non-delinquents yielded a P-value of .30, as compared with about .08 in the delinquent group. No direct comparison between delinquents and non-delinquents can be made because of the lack of uniformity in the type of records, but there is more evidence that health was a significant factor in the delinquent than in the non-delinquent group.

School grades of both delinquents and non-delinquents were analyzed in terms of trends in the interval between psychological examinations. Thirty-five of the delinquents had no school experience in this period, but of those who did there was no significant difference in the trend of their grades between the constant and improved IQ groups. In the non-delinquent group, too, the trend of grades in terms of improvement or the lack of it was no different in the improved than the constant IQ group.

Three other factors were examined in the delinquent group as possible indicators of adjustment within the institution, namely, records of discipline, work ratings, and behavior ratings. These are records kept routinely by the institution. The average number of disciplinary incidents for the constant group was 6.32 per person, and for the improved group it was 7.25 per person. Classification of types and quantity of discipline, tested by chi square, yielded a P-value between .20 and .10, which is not low enough to allow one to draw any definite conclusion about the group differences. When work ratings and behavior ratings were classified in terms of trends of improvement, lack of improvement, or poorer than before, the percentage of girls in each classification was almost identical for the constant and improved groups. Hence, it is not indicated that the group whose IQ improved showed any more improvement in adjustment to the institution than those whose IQ remained constant.

Discussion

The basic criterion for grouping these cases, of course, was the difference in the scores made on two intelligence tests. On the first test the delinquents proved to be somewhat below average, which is consistent with other reports of the intelligence level of delinquents, but the non-delinquents had a mean intelligence score close to the mean for the general population. The test-retest reliability coefficient was satisfactorily high on each group, but slightly better with the delinquents than the non-delinquents. The correlation between IQ change and time interval is so close to zero that it is impossible to say that the large changes in IQ are due to practice effect in the usual sense of the term, since that would result in a negative correlation between time interval and IQ change. Yet, 36% and 48% of each group, respectively, showed a change of more than 10 points in IQ on retest. This result, while indicating more change than that reported by many investigators, notably those using the 1916 Stanford-Binet, is by no means unprecedented. Lowell (21) with her very large group of cases reported that 71% changed 7 points or more, suggesting that the mean change must have been considerably greater than that. Brown (5) reports an average change of 6 points, which is only slightly below the average for this study, and Allan and Young (1), using the Terman-Merrill, report results identical with the ones of this study which used the Kuhlmann. They found an average gain of 8 points with 48% of the group varying more than 10 points. Variability such as these studies and the present one reveal is not simply variation around the mean inasmuch as the average change is a gain of 8 or 9 points, when computed with retention of the sign rather than change averaged regardless of sign. This gain is not to be explained by the conventional techniques of correlating IQ change with time interval or level of first IQ, as mentioned before. The range of IQ changes is quite similar whether based on the cases retested at 4 to 6 months or those retested at an interval of one year or more. The accuracy score on the Kuhlmann proves to be a rather stable index of the individual's method of working, providing he is given tests properly ranged for his level of ability, but the speed score does have a positive correlation of .41 to .44 with changes in IQ.

The achievement tests were used to find out whether those girls whose achievement level was least retarded were among those who were able to raise their IQ's on retest. As mentioned in the statement of results, the relationship of achievement to grade level is somewhat distorted by the fact that the upper ceiling of the test is grade 11.0. For this reason more significance is attached to the critical ratios than would be warranted without this factor which depressed some scores at the upper end of the distribution. With both delinquents and non-delinquents the improved group showed less retardation in achievement than those whose IQ's on retest stayed about the same. It appears that achievement level is more significant in this respect than the level of the first IQ.

The personality test which provided the richest amount of material regarding individual adjustment was the Minnesota Multiphasic Personality Inventory. Although it discriminated well between delinquents and non-delinquents, it gave no significant differences in either main group between those cases wherein the IQ improved significantly and those in which it did not. The same pattern is borne out by the comparisons of total scores on the Washburne Social Adjustment Inventory. As adjustment is measured by these tests, there is no greater maladjustment in one group than in the other.

The Terman-Miles Test of Masculinity-Feminity showed no significant differences between any groups, and the Vineland Social Maturity Scale revealed no differences in social age between those whose IQ improved and those whose IQ remained constant. The Pressey Interest-Attitude Test behaved atypically from the other tests in showing some difference between constant and improved delinquents and none between constant and improved non-delinquents. Moreover, it did not discriminate between delinquents and non-delinquents. Hence, four out of five personality tests showed no evidence whatsoever that there were marked differences in personality adjustment at the time of the first test between those whose IQ showed subsequent improvement and those whose IQ remained constant. The fifth test, the Pressey, shows such conflicting results that no conclusion may be based on it.

In general the results of the personality tests give a negative answer to the question of whether improvement in the adjustment pattern of the individual is related to improvement in IQ, although there is a slight tendency for the scores of the improved delinquents to move more toward the normal end of the distribution of scores than do the scores of the constant delinquents. With the non-delinquents, whose personality test scores were very normal on the first test, the differences in amount of score changes are not so large nor are they always in the same direction, which suggests that they are chance fluctuations. The differences with the delinquents are not great enough to be statistically reliable nor to lend real weight to the hypothesis that improvement in general adjustment results in improvement of mental test performance.

As a further attempt to find out whether changes in IQ are related to changes in other factors which are indicative of the level at which the individual is functioning, in other words his total personality adjustment, a record was made of certain other items concerning the period between tests. Studies of IQ changes frequently mention physical health as a significant factor. The health record of the delinquents does not show a reliable difference between the groups, but the critical ratio is high enough to support the belief that physical condition may be a factor with bearing on efficiency of test performance. The present results lend support to the opinion of many psychologists that a test given at a time when an individual is not physically up to par should be repeated at a more favorable time. No one can judge with accuracy just which physical ailments affect mental performance and which do not.

Returning to the questions posed at the beginning of the study, we shall sum up the results within that framework. The only differences of personality and experiential background between those adolescent girls whose IQ's show significant improvement and those whose IQ's remain relatively constant are that those who do not improve so much on the second test tend to have more retardation in level of school achievement than the variable group, and they tend to come from homes in which the parents have a lower level of education and occupation. Of the changes occurring between tests, there were no significant changes in the improved group as compared with the constant IQ group of the non-delinquents. In the delinquent group those who showed the greatest improvement in IQ were the group who had a poorer health record, so presumably it contained more girls who were not in good physical condition at the time of the first test.

This lack of proof for the significance of emotional factors in the test situation, at least for adolescent girls, carries some implications for the clinician. In recent years whenever a clinical psychologist cannot find any evident reason why an IQ changed significantly, it has become the custom to fall back on the idea that the individual may be better adjusted at the time of the second test, a statement rarely backed up by any data. Although there is a possibility that in some cases this is true, it is by no means a common or easily demonstrated cause for shifts in IQ at the adolescent or adult level. The problem in young children may be quite different and also with older adults, who often are tested at the time of a court trial or some other emotional strain. The latter situation is somewhat similar to the situation of these delinquents when given their first test, but one is not privileged to generalize these conclusions beyond the adolescent group. The great emphasis put on emotional factors by psychologists, psychiatrists, and social workers has made all persons dealing with maladjusted individuals particularly conscious of them, and the present study in no way minimizes their importance to the individual or as a determiner of behavior, but only minimizes their effect upon mental test performance.

Summary and Conclusions

Two groups of adolescent girls, 101 delinquents newly admitted to a State School for Girls and 85 public school girls, whose ages ranged from 12 to 19 and whose median grade placement was 9th grade, were given an individual intelligence test, an achievement test, and a battery of personality tests. They were retested from 4 to 15 months later. Delinquents and non-delinquents were divided into subgroups composed of those whose IQ's on retest changed 10 points or less, called the constant groups, and those whose IQ's improved more than 10 points, called the improved groups. Differences of personality adjustment were measured by computing significance of difference of the mean scores for each group on the personality tests and by classifying and measuring the significance of difference of other items, such as occupation and education of parents, racial and national background, language in the home, work experience, length of time out of school, type of delinquency, other known social problems in the family, and home ratings. In the interval between tests school grades and health records were compared, and—in the delinquent groups—discipline records, and work and behavior ratings. Differences were computed between the mean amount of change in scores on personality tests given in the second examination.

The following conclusions appear to be warranted:

1. The non-delinquents showed as large shifts in IQ on retest as the delinquents.

2. In both delinquent and non-delinquent groups those who showed less improvement in IQ on retest showed the greater amount of achievement retardation.

3. There were no significant differences in the degree of adjustment at the time of the first test or in the experiential background between those who improved on retest and those whose IQ remained relatively constant.

4. There were no significant differences between the constant and improved IQ groups in relation to the personal-social history which took place in the interval between tests.

5. There were no significant differences of personality adjustment between the constant and improved IQ groups at the time of the second test.

6. Concomitant personality factors, as measured by these tests and the criteria set up by this study, do not have any demonstrable effect on the changes in mental test performance between test and retest of adolescent girls.

Received June 12, 1944.

References

1. Allan, M. E., and Young, F. M. The constancy of the intelligence quotient as indicated by retests of 130 children. *J. appl. Psychol.*, 1943, 27, 41–60.
2. Baldwin, B. T. Child Psychology, a review of the Literature, Jan. 1, 1923 to March 31, 1928. *Psychol. Bull.*, 1928, 25, 629–697.
3. Baldwin, B. T., Stecher, L. I., and Smith, M. General reviews: mental development of children. *Psychol. Bull.*, 1923, 20, 665–683.
4. Bronner, A. F. Attitude as it affects the performance of tests. *Psychol. Rev.*, 1916, 23, 300–331.
5. Brown, A. W. Changes in intelligence quotients of behavior problem children. *J. educ. Psychol.*, 1930, 21, 341–350.
6. *Burks, Barbara* S. A summary of literature on the determiners of the intelligence quotient and the educational quotient. *Yearb. Nat. Soc. Stud. Educ.*, 1928, 27 (II), 319–325.

7. Dickson, V. E. *Mental tests and the classroom teacher.* New York: World Book Co., 1923 (Esp. pp. 69–71).
8. Doll, E. A. A genetic scale of social maturity. *Amer. J. Orthopsychiat.*, 1935, 5, 180–188.
9. Foran, T. G. The constancy of the intelligence quotient; a review. *Cath. Univ., Amer. educ. Res. Bull.*, 1927, **1**, No. 10.
10. Foran, T. G. A supplementary review of the constancy of the intelligence quotient. *Cath. Univ. Amer. educ. Res. Bull.*, 1929, **4**, No. 9.
11. Ford, C. A. The variability of IQ's for psychopaths retested within fifteen days. *Psychol. Clin.*, 1929, 18, 199–204.
12. Gildea, H., and Macoubrey, C. Factors affecting the constancy of the intelligence quotients of problem children. *Smith Coll. Stud. Soc. Work*, 1933, 3, 229–247.
13. Hallowell, Dorothy K. Validity of mental tests for young children. *J. gen. Psychol.*, 1941, **58**, 265–288.
14. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *J. Psychol.*, 1940, 10, 249–254.
15. Hildreth, Gertrude. Occupational status and intelligence. *Person. J.*, 1934, 13, 153–157.
16. Jewett, S. P., and Blanchard, Phyllis. Influence of affective disturbances on response to the Stanford-Binet test. *Ment. Hyg.*, N. Y., 1922, 6, 39–56.
17. Kelley, T. L., Ruch, G. M., and Terman, L. M. *Stanford Achievement Tests: Manual of Directions.* Yonkers-on-Hudson, New York: World Book Company, 1940.
18. Kuhlmann, F. *Tests of mental development.* Minneapolis, Minn.: Educ. Test Bureau, 1939.
19. Leahy, S. R., and Fox, E. J. An investigation of the effect of the emotional factor on the intelligence quotient. *J. juv. Res.*, 1930, 14, 260–266.
20. Lincoln, E. A. The constancy of the I.Q. *J. educ. Psychol.*, 1922, 13, 484–495.
21. Lowell, Frances E. A study of the variability of I.Q.'s in retest. *J. appl. Psychol.*, 1941, 25, 341–356.
22. Matthew, J., and Luckey, B. Notes on factors that may alter the intelligence quotient in successive examinations. *Yearb. Nat. Soc. Stud. Educ.*, 1928, 27 (I), 411–419.
23. Miller, E. A. Emotional factors in intellectual retardation. *J. ment. Sci.*, 1933. 79, 614–625.
24. Nemzek, C. L. The constancy of the I.Q. *Psychol. Bull.*, 1933, **30**, 143–167.
25. O'Neill, H. O. Variations in the intelligence quotients of 105 children. *Child Develpm.*, 1937, **8**, 357–363.
26. Pressey, S. L., and Pressey, L. C. Development of the interest-attitude tests, *J. appl. Psychol.*, 1933, 17, 1–16.
27. Richards, T. W. Individual variations in I.Q. and analysis of concomitant factors. *Psychol. Bull.*, 1940, 37, 442–443.
28. Schott, E. L. Variability of mental ratings in retests of neuropsychiatric cases. *Amer. J. Psychiat.*, 1930, 10, 213–227.
29. Taussig, F. *Principles of economics* (2nd rev. Ed.). New York: Macmillan, 1920.
30. Terman, L. M., and Miles, C. C. *Sex and personality.* New York: McGraw-Hill, 1936.
31. Thorndike, R. L. Constancy of the I.Q. *Psychol. Bull.*, 1940, 37, 167–186.
32. Wallin, J. E. W. Results of retests by means of the Binet Scale. *J. educ. Psychol.*, 1921, 12, 392–400.
33. Washburne, J. N. A test of social adjustment. *J. appl. Psychol.*, 1935, 19, 125–144.
34. Williams, J. H. The Whittier scale for grading home conditions. *J. Delinqu.*, 1916, 1, 273–286.

Studies in the Symbolism of Voice and Action: V. The Use of Behavioral and Tonal Symbols as Tests of Speaking Achievement

Franklin H. Knower

University of Iowa

There are two general lines of procedure which may be followed in initiating a program of research in test construction. First, one may begin with a large number of items selected because of some presumed diagnostic value in the test, and then, through a program of item analysis, select and retain those items which contribute most to the value of the test. In following the second method, one begins with a limited number of items, retains those found to be useful and then adds to them as the evidence indicates a need for additional items to improve the value of the test. The first method may be said to be most satisfactory in dealing with tests of a nature comparable to those of already known value. The second method is to be preferred in the development of tests which are essentially different from those of established patterns, for unless a test of a limited number of items can be shown to have some usefulness, the possibility that a test of a greater number of items will be of value is immediately open to question.

The second method just described has been employed in the work on the tests used in this project. The data to be presented are to be interpreted, therefore, as preliminary findings.

The use of tonal and behavioral symbols as a supplement to linguistic means of communication long have been considered by critics as significant factors in the process of speaking. A number of recent studies have supported this impression.

Monroe[1] found that two of the more important elements in the second factor revealed by his factor analysis of the characteristics of good speech were "Used good gestures" and "Voice not monotonous." Barnes[2] reports "Monotonous voices" and "Poor bodily control" as relatively high frequency faults among the speech students he studied. In a study by

---

[1] Monroe, Alan H. The measurement and analysis of audience reaction to student speakers. *Studies in Higher Education*, Purdue University, xxxii, 1937.

[2] Barnes, Harry G. A study of the speech needs and abilities of students in a first course in speech training at the college level. Unpublished Doctoral Dissertation, State University of Iowa, 1932.

Gilkinson and Knower[3] "Monotonous voices," "Inanimate bodies" and "Little facial expression" were not only indicated as high frequency faults, but were also shown to be speech characteristics clearly differentiating speakers of superior from inferior quality. At one point in their report on the Michigan Cooperative Study, Hayworth and colleagues[4] have indicated that the number of "Meaningful facial expressions per minute" was found to be a factor of relatively high weight in the determination of "Total effectiveness of speech delivery." At another point in their report on the evaluation of specific tests of "Vocal interpretative ability," "Pantomiming ability," and "Facial expression," their data indicate moderate to marked relationship with "Public speaking effectiveness" for the first two tests; and approximately a zero correlation for the test of "Facial expression." The absence of an indication of relationship in the second test may be attributed to the inadequacy of the particular test used.

With the exception of part of the data in the Hayworth study, all of the data obtained on this problem in previous studies have been secured by the processes of rating a speaker on his use of these symbolic processes during the activity of speaking. The data in this study, on the contrary, are not derived from an evaluative judgment on the quality of the performance, but as explained later are actual interpretations of meaning. This investigation was undertaken to secure data which might throw light on the answers to the following questions.

1. Is it possible to develop reliable and valid objective group tests of a speaker's skill in the use of tonal and behavioral symbolism as such?

2. What can we learn from such objective tests about the relationships of such skills to the total effectiveness of the speaker in speaking, and to other characteristics of the speaker as a person?

The tests upon which this study are based are adaptations to group testing purposes of the measuring instruments used in the study of tonal and visible symbolism reported in the *Quarterly Journal of Speech* by Dusenbury and Knower.[5] The test form consists of a sheet of paper containing in the left hand margin a list of eleven emotional states each designated by three terms selected to facilitate the recognition of the particular qualitative or quantitative differences in the emotional states to be expressed. A series of columns across the sheet are arranged to permit judges to record their interpretations of each performer's stimulation in each performance. As each student is tested his name is placed at the head of a column and recorders' interpretations are placed below it. The recorder's name and other data are provided at the top of the judging sheet.

---

[3] Gilkinson, Howard, and Knower, Franklin H. Psychological studies of individual differences among students of speech. Univ. of Minnesota, Department of Speech, 1930.

[4] Hayworth, Donald. *A research into the teaching of public speaking.* National Association of Teachers of Speech, Detroit, 1940. Pp. 231.

[5] Dusenbury, Delwin, and Knower, Franklin H. Experimental studies of the symbolism of action and voice. *Quart. J. Speech*, 1938, 24, 424–436; 1939, 25, 67–74.

Directions to the performers for the test of behavioral symbolism were as follows: "At the next meeting of the class you will be called upon to indicate the way you would express the eleven emotional states indicated on this sheet by *facial* gestures or pantomime. Although you are expected to depend primarily upon facial expression you may allow your body to adapt to the expression you are trying to indicate in the face. You are requested especially to avoid hand gestures of conventional meanings. You are to use your articulators as if you were saying the letters of the alphabet from 'a' to 'k,' but you are not to form words with your lips which might be interpreted by lip readers. You are asked simply to express the emotions indicated as you would simulate them if you were portraying the emotions of a character in a story you were telling or in a play you were acting. During the class period I will call you to the desk on the platform one by one and hand you a series of cards one at a time on which have been typed the terms for the eleven emotional states to be expressed. As I hand out each card I will call out numbers from '1' to '11' and record these numbers on a key sheet for correcting the responses of the observers. The stimulus cards will be shuffled after each performance to vary the order for the next performance. You are expected to take only two to three minutes for the entire series of eleven expressions and you must therefore respond as quickly to the cue cards as you would to the cue lines in a play."

The directions for the members of the class, who served as judges when not performing, were as follows: "You are to place each performer's name at the top of the column in which you record the judgments of his performance. As I call each number observe the speaker carefully, look quickly down the list of eleven emotional states and record the number in the proper column opposite that emotional state which in your judgment the speaker is simulating. You are to record a judgment for each number, and record each number opposite only one set of terms. After the first expression for each performer, if you interpret a later expression as the simulation of an emotion for which you already have one judgment, you may place the second number also opposite the terms for that emotion. For every set of terms on which you record more than one judgment, you will leave one set of terms without a number. You should familiarize yourselves with the list in order to record your judgments rapidly."

The directions for performance in the tonal test differed from those for the visible symbolism only in the following manner. "At the next meeting of the class I will call each of you to the back of the room and ask you to simulate a tonal expression of each of the emotional states indicated on this sheet. You are to use no words but in giving tonal body and pattern to the expression you are to articulate the letters of the alphabet, from 'a' to 'k.' "

The process of judging the tonal expressions differed from the process of judging the facial expressions only in that the judges listened to the tonal expressions rather than observed the behavior of the speakers. As has been indicated, the instructor in the class made a key during each performance for the purpose of checking the correctness of interpretations. Since in no case were there fewer than sixteen persons in the class section tested, there were at least fifteen judges for each performance.

The tests were scored in two ways. First the number of correct judgments for each performer were divided by the total number of judgments when checked against the key to provide a score in terms of a percentage of audience comprehension of each performance. Secondly, the number of correct judgments rendered by each judge was divided by the total number of his judgments to provide a score in terms of the percentage of his comprehension of the performances of all other members of the class. Although the class average for the performances and judgments were of course always the same, marked differences in individual ability of both types occurred within most classes.

Before analyzing the data from the study, it may be well to consider some of the advantages and limitations of tests of this type. On the question of their desirable features, I wish to call attention first to the fact that these tests are actual tests of the social dynamics of speech. The score in no way depends upon an auditor's interpretative judgment of the excellence of the performance, but only on the specific percentage of the speaker's intelligibility. Since the tests are of this type they may be used to analyze individual differences in the nature of the behavior of both speakers and auditors. A relatively large number of persons can be tested in the average class period in a situation which approaches normal student speaker-auditor relationships. The fact that the test may function as a learning activity commends it as a classroom exercise apart from its test features.

Such tests may be criticized as abstractions in that they isolate tonal and behavioral symbolism from their normal accompaniment in speech of linguistic expression of ideas. While this criticism is obviously sound, it is probably not more true for the test in question than for typical tests of articulatory or linguistic skill. Although intelligent persons do not ordinarily go about engaging in unprovoked facial activity or vocal noises, neither do any but school teachers and other intellectually curious persons go about randomly articulating phonemes, declining adjectives or parsing sentences. These characteristics of tests probably are necessary limitations of the problem at hand. Although the tests may not provide an adequate index of refinement in the use of non-linguistic symbolism for advanced students, the limitation of scores to intelligibility provides a sufficient top for most students with limited speech skills. A third question may be raised concerning a test of a speaker's performance which limits the index of his skill to the comprehension ability of his audience. This might be a serious limitation were it not for the fact that in almost every class tested the range of scores on performance has been moderately high. Data will be presented which indicates that in a class of at least fourteen judges the reliability of the tests is sufficiently high for purposes of group analysis.

The major data on the project are presented in Table 1. The correlation of the performance scores of subjects as interpreted by seven judges with scores as interpreted by seven additional judges produces what may be called a split-half index of the reliability with which performances are judged. When corrected for the split-half technique, the reliability correlations for scores on behavioral symbolism were +.93 and for tonal symbolism +.87. The individual performances, then, were reliably interpreted. To determine whether or not the expressional skill of the individual was consistently characteristic of his performance, each test was repeated after an interval of one week for 100 subjects. The second line of the table indicates these test-retest correlations. Although these correlations indicate a substantial amount of consistency in the traits measured, they also indicate that there was considerable variation in performance on the second test, with skill in the use of tonal symbolism varying less than skill in the use of behavioral symbolism. Since the indices or scores for each performer are based on the average accuracy of interpretation of fifteen listeners,—where fourteen judges produced reliability of +.93,—these correlations cannot be attributed to the unreliability of the particular scores. These test-retest correlations were seriously affected by the average gain in the effectiveness of performance on the second test, and by the fact that the range of scores was considerably reduced on the second test. Since the average gain in effectiveness of the use of tonal symbolism was as high as the average gain in effectiveness of the use of behavioral symbolism the higher test-retest correlations for the test of tonal symbolism indicate that fewer subjects approximated the top of the tonal test than the behavioral test. This assumption is sup-

Table 1

Reliability and Validity Correlations

	Behavioral Symbols		Tonal Symbols	
	Perf'nce	Judg'nt	Perf'nce	Judg'nt
Seven with Seven Judges (Corrected)	.93 ± .01		.87 ± .02	
Test-Retest	.52 ± .05	.69 ± .03	.66 ± .03	.92 ± .01
Speech Rtgs. for General Effect'ness	.47 ± .04	.56 ± .04	.25 ± .05	.56 ± .04
Speech Rtgs. for Adjustment	.42 ± .05	.40 ± .05	.34 ± .05	.41 ± .05
Speech Rtgs. on Phonation			.31 ± .05	.46 ± .04
Performance with Judgment (1st Test)		.57 ± .03		.49 ± .04
Performance with Judgment (2nd Test)		.33 ± .06		.55 ± .06

ported by the fact that the mean level of performance in the use of tonal symbols was considerably lower than the mean level of the skills in behavioral symbolism.

The correlation indices of validity may be best described as moderate to marked in comparison with other tests of this type. These data are inconsistent with those reported by Hayworth in that this test of performance in the use of facial expression appears to be more closely related to speech skill than does the test of the use of tonal symbolism. An interesting feature of these correlations is found in the fact that ratings on performance are more closely correlated with skill in judgments of performance than with skill in performance itself. If this phenomenon should be found to be consistently true of this type of test, and if these validity correlations may be improved by the development of the tests, it will mean that since judgment scores may be obtained more conveniently than performance scores, the potential usefulness of the tests will be greatly improved.

Table 2

Means and Standard Deviations of the Distributions

	First Performance			Second Performance		
	Means	S.D. of Perf.	S.D. of Judg.	Means	S.D. of Perf.	S.D. of Judg.
Behavioral Symbols:						
Freshmen (Required)	62.70	10.40	14.10	71.10	13.10	10.50
Sophomores (Elective)	70.90	10.85				
Tonal Symbols:						
Freshmen (Required)	45.10	13.70	9.15	57.40	13.50	10.90
Sophomores (Elective)	66.80	13.95				
Acting Class	83.30	4.95				

Table 2 contains some additional data on validity in the form of means and standard deviations of distributions of scores for various groups. A group of 50 sophomores in an elective course in speech were compared with 100 freshmen in a required course and found to be eight per cent more intelligible in the use of visible symbolism and twenty-one per cent intelligible in the use of tonal symbolism than were the freshmen. The difference between the two groups is highly significant in the use of tonal symbolism and probably reliable in the use of visible symbolism. Nineteen students in an advanced class in acting received scores on the tonal test that were significantly higher than the scores received by the sophomores.

The standard deviation of the judgment scores of the freshmen on the first tests are here seen to be considerably higher than their deviation on the second tests. The amount of improvement on the second tests is also indicated here in terms of higher mean scores on the second tests. I should like to point out that, although these mean scores are considerably lower than those reported by Dusenbury and Knower's advanced performers, they are still high enough to indicate that a highly significant amount of communication may take place through the use of tonal and behavioral symbolism by the average college student.

In concluding, I wish to comment on the last two lines of correlations in Table 1. In spite of apparent differences in skill in performance and judgment these correlations indicate a considerable relationship between these two traits. The relationship appears to be more constant in the area of tonal symbolism than in the area of visible symbolism. These data support the suggestion previously advanced that a group test of skill in the judgment of tonal and behavioral symbolism might be developed which will provide a useful test of skill in performance. Such tests are useful as classroom checks on the development of significant aspects of speech achievement.

Received May 11, 1944.

Participation in High-School Football as a Factor Affecting College Attendance and Scholarship

Erwin J. Henning and Harold D. Carter *

School of Education, University of California

Does participation in high-school football exert a significant influence upon the educational plans and later educational careers of high-school students? If so, what are the effects of such participation, and how do they operate? These questions, which have long been subjects for heated discussion as well as for organized inquiry, are approached again here, with a variation in technique that appears to be new to this field of investigation.

In order to develop a standard for estimating the effect of playing football on the educational careers of the players, a control group was also studied. Each high-school football player was matched with a classmate who did not play football. The college attendance and achievement of the control group were used as norms for comparison with the performance of the athletes.

Records were taken for 2,875 pupils, including 220 football athletes, graduating from four large high schools in the years 1935–1937 inclusive. These high schools were all situated in the urban area on the eastern side of San Francisco Bay. The 220 football players were matched with 220 of their classmates on the basis of six criteria, namely: (1) school; (2) date of graduation; (3) high-school scholarship; (4) average measured intelligence; (5) reading quotients as measured by standardized tests; and (6) college preparatory study load, as measured by the number of units earned in courses which satisfy prescribed college entrance requirements. In the pages which follow, a detailed report is made concerning the high-school work and the college attendance and scholarship of the two groups.

Significant Earlier Studies

The controversy over college recruiting and subsidizing of athletes, which was at its height in the late 1920's, precipitated the famous Carnegie investigation reported in 1929 in the bulletin on American College Athletics (4). In that research the case study technique was emphasized.

---

* The writers are indebted to Dr. Robert Gordon Sproul, President of the University of California, for his encouragement, and for financial assistance which made the study possible.

Investigators traveled from college to college studying conditions pertaining to the treatment of athletes, and reporting their findings. An accompanying survey of the literature resulted in a separate publication reported in the 24th bulletin of the society (5). This is an authoritatively comprehensive survey of the literature on athletics up to 1929. By 1933, the Carnegie Foundation had spent $125,000 on its investigations of athletics.

The Carnegie research found previous studies on the relation between scholarship and athletics ill-controlled and inconclusive. Lack of uniform definitions made it impossible to compare results in the various schools. For this reason, Howard Savage, who directed the Carnegie investigation, sponsored a study of scholarship and athletics at Columbia University. This study served as a model for coordinated projects in 52 colleges (15). Nearly all later research on the scholarship of athletes in college has also followed Savage's model. The various investigations demonstrate that the athlete tends to be slightly inferior to the non-athlete in measured intelligence and in grades earned, but not enough so as to interfere seriously with his scholastic work. Athletes take a normal study load and a normal variety of college courses, and have slightly smaller academic mortality than non-athletes; however, more athletes than non-athletes receive grades near the failing mark, and more are on probation. These studies show a difference among sports, with sports for individuals, such as tennis, golf, track, etc., ranking high when judged by the academic achievement of participants, whereas team, spectator sports such as baseball and football rank low when judged by the same criteria. Later college investigations such as those by Tuttle (18), Hackensmith and Meller (10), Maney (12), and others tend to confirm the findings reported in the Carnegie Foundation studies.

In the field of high-school athletics, Reals and Rees (14), Mathews (13), Hull (11), and Allen (1) find athletes slightly inferior to their non-athletic fellows. In no case is the reported difference large. Cormany (7), Cook (6); Beu (2), Schulman (16), and Shannon (17) report high-school athletes equal or slightly superior to their classmates. Several of these reports show that while athletes tend to be inferior to their fellow students who are active in other extra-curricular activities, they tend to be superior to those who have no extra-curricular interests. Davis and Cooper (8) reviewed 41 studies in this field, and concluded that probably the non-athletes are slightly superior to the athletes, but that the difference is not significant either statistically or educationally.

The above are but a sampling of the reports on athletes in high school and in college. Very few studies have followed high-school athletes into college. To do so seems desirable, since the controversy rages over the inducing of high school athletes into college by promises of glory, financial rewards, or lowering of academic standards. Buck (3) studying 1098 high-school boys in Colorado, found that 21.8 per cent of the athletes planned to go to college, whereas only 11.6 per cent of the non-athletes planned to do so. He gave no data as to the relative college qualifications of the two groups. Among the athletes, the athletic prestige of the college stood first in determining a specific choice, whereas this was last in the list for non-athletes. Relatively more athletes than non-athletes persisted in college until they received a degree. Here again, only meager data are presented as to the relative qualifications of the two groups.

The present study is unique in that it begins with high-school groups of equivalent academic qualifications, and follows them into college.

High-School Records

A by-product of matching football athletes with controls was the accumulation of a large mass of data regarding the high-school records of all athletes and non-athletes. These data are entirely in agreement with the literature on the scholarship and ability of high-school athletes. Table 1 presents the results. They show the non-athletes to be slightly though consistently superior to the football players.

These data have more value than merely substantiating the extensive literature in the field. By falling completely in agreement with numerous other investigations, they demonstrate the typicalness of the high-school sample used here, and hence indicate the validity of our further findings. They encourage one to generalize with confidence beyond the four specific high-school populations studied.

The above-mentioned data are for total populations, showing that the football athletes closely resemble the rest of the student body. For the matched pairs, however, the resemblance is still closer as a result of the process of matching. The median difference between the football players and their controls in grade point ratio, intelligence quotients, reading quotients, and type of high-school load is only .19 times as large as its standard error, demonstrating the practical identity of the qualifications of the matched groups. This identity is necessary to satisfy the basic premise of this study. If athletes and their controls are substantially identical in the qualifications which ordinarily determine college entrance and success, then the average college experiences of the two groups should also be identical within statistical limits of chance variation, provided that playing football in high-school has no influence in college. Conversely, any real variation between the experiences of the two groups is attributed to the football qualifications of the experimental group.

Table 1

High School Record of Football Athletes Compared with That of Various Other Groups

	N	Mean	Comparing Lines	Diff.	$\frac{\text{Diff.}}{\text{S.E.}_{\text{Diff.}}}$
Grade Point Ratio:					
1. All football players	220	2.47			
2. All others	2655	2.54	1 & 2	.07	1.82
3. Matched control group	220	2.47	1 & 3	.00	.00
4. Athletes attending college	133	2.62			
5. Controls attending college	99	2.76	4 & 5	.14	2.10
Intelligence Quotients:					
1. All football players	220	106.44			
2. All others	2655	107.94	1 & 2	1.50	1.72
3. Matched control group	220	106.80	1 & 3	.36	.32
4. Athletes attending college	133	110.31			
5. Controls attending college	99	112.71	4 & 5	2.40	1.92
Reading Quotients:					
1. All football players	220	107.10			
2. All others	2655	110.58	1 & 2	3.48	3.07
3. Matched control group	220	108.84	1 & 3	1.54	1.05
4. Athletes attending college	133	110.42			
5. Controls attending college	99	113.90	4 & 5	3.48	2.06
Number of Half-Units College Preparatory Courses:					
1. All football players	220	25.98			
2. All others	2655	25.30	1 & 2	.67	1.69
3. Matched control group	220	25.57	1 & 3	.40	.79
4. Athletes attending college	133	27.95			
5. Controls attending college	99	28.13	4 & 5	.17	.29

Plans to Attend College

More football athletes than controls planned to go to college, as indicated by transcripts sent. Table 2 shows that 28.2 per cent more athletes than controls sent transcripts to college. This difference is 7.52 times its standard error, and hence is statistically significant. The University of California received more enquiries from athletes than from controls, but the difference is not significant. All colleges receiving fewer than three enquiries from prospective students in this group were classified as "miscellaneous" colleges. In the aggregate, these miscellaneous colleges received 47 enquiries from football students and 37 from controls. The difference is not significant. Besides the University of California and the miscellaneous colleges, thirteen other colleges received transcripts from prospective students. Of these, the seven receiving the majority of the transcripts were all local institutions which emphasize football. This is the group of colleges which received significantly more transcripts from football athletes than from members of the control group.

Table 2

Planned and Actual College Attendance of Members of Matched Groups

	Football Group	Control Group	Diff.	$S.E._D$	$\frac{\text{Diff.}}{S.E._{Diff.}}$
Planned college attendance as indicated by transcripts:					
Total number sending transcripts	201	139			
% of 220 sending transcripts	91.4	63.2	28.2	3.75	7.52
Number planning to go to U.C.	77	66			
% of 220 planning to go to U.C.	35.0	30.0	5.0	4.46	1.12
Number planning to go to 13 colleges	78	36			
% of 220 planning to go to 13 colleges	35.4	16.4	19.0	4.07	4.67
Actual college attendance of members of matched groups:					
Total number going to college	133	99			
% of 220 going to college	60.5	45.0	15.5	4.69	3.30
Number going to U.C.	62	56			
% of 220 going to U.C.	28.2	25.5	2.7	4.21	.64
Number going to 13 colleges	53	24			
% of 220 going to 13 colleges	24.1	10.9	13.2	3.54	3.73

College Attendance and High-School Scholorship

Table 2 summarizes facts from the college attendance records of members of the matched groups. 2.8 per cent more athletes than controls attended the University of California; the difference is negligible. 13.3 per cent more athletes than controls attended the thirteen colleges; this difference is significant. Eighteen football athletes and nineteen controls attended miscellaneous colleges. Thus a total of 133 athletes attended college, compared with 99 members of the control group. The difference is 3.09 times as large as its standard error. The excess college attendance of athletes over controls is due almost entirely to the fact that the thirteen colleges accepted more athletes than members of the control group. The inference follows that the excess college attendance of the athletes is due to their high-school football experience.

The data in Table 1 permit comparison of the qualifications of the athletes who attended college with the qualifications of the controls who did so. The athletes tend to be inferior to the controls. Which colleges *accepted these inferior students?* The qualifications of the members of the matched groups who attended the miscellaneous colleges are nearly equivalent. The data for members of the matched groups attending the thirteen colleges and the University of California are compared in Table 3. At the University of California the high-school football students who entered college came with generally better qualifications than those of their controls. At the thirteen colleges the reverse was true. Table 3 emphasizes this situation by comparing the football students who attended the University of California with the football students who attended the thirteen colleges. A similar comparison is made for members of the control group who attended these schools.

Table 3

Comparison of High-School Qualifications of Football Athletes and Control Groups Attending the University of California and the 13 Colleges

	Football Groups		Control Groups	
	U.C.	13 Colleges	U.C.	13 Colleges
Number attending	62	53	56	24
Mean high-school grade-point ratio	2.90	2.38	2.89	2.58
Difference between means	.52		.31	
Standard error of difference	.08		.13	
Critical ratio	6.23		2.37	
Mean high-school IQ	114.07	107.00	113.43	110.33
Difference between means	7.07		3.10	
Standard error of difference	1.74		2.29	
Critical ratio	4.06		1.35	
Mean high-school reading quotient	112.10	108.98	113.11	113.01
Difference between means	3.12		.10	
Standard error of difference	2.46		2.82	
Critical ratio	1.27		.04	
Mean number half-units college preparatory courses	29.71	26.25	28.82	27.38
Difference between means	3.46		1.44	
Standard error of difference	.93		.60	
Critical ratio	3.72		2.40	

In all cases the high-school qualifications of students attending the University of California were superior to the average qualifications of those who attended the thirteen colleges. This is true both for athletes and for controls. The University of California attracts the better students from this area. However, the differences between the qualifications of the controls who attended the University of California and the thirteen colleges were relatively small and statistically non-significant. The critical ratios are displayed in Table 3. Thus while the state university attracted generally better-qualified students than the thirteen colleges, the differences were small for ordinary students. The seven local colleges emphasizing football, however, attracted and accepted football students who were definitely inferior in academic ability and achievement. Since this is not true for non-football students, the results suggest a lowering of standards by these local colleges, for prospective football players. The same results indicate that at the University of California standards were not lowered for football players.

Table 4

Average Grade-Points Ratios Earned in College by Members of Matched Groups

	N	Mean	Diff.	S.E._{D}	$\frac{\text{Diff.}}{\text{S.E.}_{\text{Diff.}}}$
University of California					
Football Group	62	1.272			
Control Group	56	1.141	.131	.081	1.62
Thirteen Colleges					
Football Group	53	0.782			
Control Group	24	1.105	.323	.187	1.73
Totals					
Football Group	115	1.046			
Control Group	80	1.133	.087	.086	1.01

Table 4 shows that at the University of California the football athletes earned slightly higher grades than members of the control group The difference is so small, however, as to be insignificant. On the other hand, the grades received by football players in the thirteen colleges were inferior to those of their controls. These results are exactly in agreement with the high-school qualifications of the groups concerned. Thus while the thirteen colleges accepted football players with inferior qualifications, the grades given them were not out of line with their abilities. Apparently the influence which gets these inferior students into college does not extend to the classroom teachers who assign them grades.

Other Comparisons

A number of additional comparisons are furnished in Table 5. Of the football players entering college, 45.1 per cent were eventually graduated, as compared with 42.4 per cent of the controls. Thus, in spite of inferior entering qualifications, the athletes were graduated at least as frequently as were the controls.

The types of college courses pursued by athletes and controls were equally distributed. While the athletes took fewer technical and com-

Table 5

Miscellaneous Data Concerning Members of Matched Groups Who Attended College

	Football Group		Control Group	
	N	%	*N*	%
University of California:				
Number who attended	62		56	
Average no. of semesters	7.6		7.0	
Number graduating	36	58	33	59
Semesters on probation	53	11.3	33	8.4
Number dismissed	10	16.1	7	12.5
Thirteen Colleges:				
Number who attended	53		24	
Average no. of semesters	4.3		4.0	
Number graduating	10	25.8	7	29.2
Semesters on probation	25	10.9	13	13.5
Number dismissed	7	13.2	1	4.2
Miscellaneous Colleges:				
Number who attended	18		19	
Average no. of semesters	4.3		3.7	
Number graduating	5	37.8	2	10.5
Total Group:				
Number in college	133		99	
Average no. of semesters	5.8		5.6	
Number graduating	60	45.1	42	42.4
Semesters on probation	78	11.2	46	9.4
Number dismissed	17	14.7	8	10.0

mercial courses than the controls, they took about the same number of professional courses, arts courses, and vocational courses. The distribution is normal enough to refute the argument that athletes take only easy courses in college.

As to the relative numbers of the two groups who were on probation, numbers of honors received, and numbers of semesters in attendance, the data tend rather consistently to favor the control group, but by very small margins.

Star Athletes

From the group of 220 athletes, the sixty most outstanding football players were selected. These were compared with the other athletes, with the controls, and with the general population. No data, either in high-school or in college, yielded sufficiently large differences to permit one to conclude that the sixty star athletes enjoyed an experience in any way atypical for the whole football group.

Summary and Conclusions

A study has been made comparing the college attendance and scholarship of 220 high-school football athletes with similar data for 220 of their classmates who did not play football. The two groups were carefully matched for intelligence and high-school scholarship. The data lead to the following conclusions:

1. In comparison with other students of equivalent academic qualifications, football athletes more often plan to go to college.
2. A greater proportion of football athletes actually do go to college.
3. A greater proportion of football athletes graduate from college.
4. The football athletes have slightly inferior high-school records.
5. The football athletes tend to enter certain colleges in which football is a prominent sport. It is inferred that standards of admission in these colleges are lowered for prospective football players.

Received May 26, 1944.

References

1. Allen, J. Houston. A study of the scholastic records of high school football squad members and high school non-athletes. Master's thesis, 1935, Southern Methodist University, 89 pages.
2. Beu, F. A. The mental ability of athletes in comparison with non-athletes in high school. *Amer. Sch. Board J.*, 1926, 73, 155.
3. Buck, J. Raymond. High school athletic participation and planned college attendance. Master's thesis, 1936, Colorado State College.
4. Carnegie Foundation for the Advancement of Teaching: *American college athletics.* Bull. No. 23, New York; 1929.
5. Carnegie Foundation for the Advancement of Teaching: *The literature of american school and college athletics.* Bull. No. 24, New York: 1929.
6. Cook, William A., and Thompson, Mabel. Comparisons of letter boys and non-letter boys in a city high school. *Sch. Rev.*, 1928, 36, 250–258.
7. Cormany, W. J. B. High school athletes and scholarship as measured by achievement tests. *Sch. Rev.*, 1935, 43, 456–461.
8. Davis, Elwood Craig, and Cooper, John A. A resume of studies comparing scholarship and abilities of athletes and non-athletes. *Res. Quart. Amer. phys. Educ. Ass.*, 1934, 5, 68–78.
9. Eaton, Dorothy, and Shannon, J. R. College careers of high school athletes and non-athletes. *Sch. Rev.*, 1934, 42, 356–361.
10. Hackensmith, C. W., and Meller, L. A. A comparison of the academic grades and intelligence scores of participants in intermural athletics at the University of Kentucky. *Res. Quart. Amer. Ass. health & phys. Educ.*, 1938, 9, 94–99.
11. Hull, J. D. A comparison of the grades of athletes and non-athletes. *Amer. Sch. Board J.*, 1924, 69, 107–109.
12. Maney, C. A. The grades of college football students. *Sch. & Soc.*, 1933, 38, 307–308.
13. Mathews, Steve. A comparative study of intelligence, attitudes, and ratings of high school athletes and non-athletes. Master's thesis, 1938. East Texas State Teacher's College, 43 pages.

14. Reals, W. H., and Rees, R. G. High school lettermen—their intelligence and scholarship. *Sch. Rev.*, 1939, 47, 534–539.
15. Savage, Howard J. *College athletics and scholarship.* Carnegie Foundation for the Advancement of Teaching, twenty-second annual report, 1927, pages 49-65.
16. Schulman, Herman. A study of the scholarship of students participating in extracurricular activities, with special reference to athletics. *Bull. High Points*, 1930, 12, 3–7.
17. Shannon, J. R. Scores in English of high school athletes and non-athletes. *Sch. Rev.*, 1938, 46, 128–130.
18. Tuttle, W. W., and Beebee, F. S. A study of the scholastic attainments of letter winners at the State University of Iowa. *Res. Quart. Amer. Ass. health, phys. Educ. & Recr.*, 1941, 12, 174–180.

Two Methods of Combining Attitudes of Like, Indifference and Dislike Into One Score

Philip Eisenberg

Columbia Broadcasting System, Inc., New York City

In radio research, as in many other fields of investigation, the social scientist frequently asks his subjects to express their like or dislike of a stimulus. Since it is undesirable to force the individual to express like or dislike when he is not sure of his opinion or when he does not feel strongly either way, he is frequently permitted to express his doubt or indifference.

Thus, the listeners' judgments can be completely distributed among the three categories of like, indifference and dislike. For convenient comparisons of reactions to different stimuli, it is desirable to express the three percentages in one combined score. Two such combined scores have been used. The problem of this investigation is to determine the relative merits of the two scores as applied to the attitudes of groups of listeners responding to various radio programs.[1]

The Two Scores

A. The LD-score. The customary technique of combining the three reactions into one score is to assign relative weights to each reaction and sum the resultant products. One way of doing this is to subtract percentage dislike from percentage like. Minus signs can be eliminated by assigning weights of 2 for like, 1 for indifference, and 0 for dislike. Either method yields the same result. We will refer to this score as the Like-minus-Dislike score, or for convenience, as the LD-score.

B. The S-score. Lazarsfeld and Robinson[2] have proposed another technique of combining the three reactions into one score, which is essentially a sigma score, and can therefore be called an S-score. It is based on three assumptions:

---

[1] All data were obtained with the use of the Lazarsfeld-Stanton program analyzer technique as used by the Program Analysis Division of the Columbia Broadcasting System. In this technique, the listener is asked to express his attitudes throughout a broadcast by pressing a green button when he likes what he is listening to, a red button when he dislikes it, and neither button when he is indifferent. The pressing of the buttons is automatically recorded on a moving tape. For a more detailed description of the program analyzer, see the article by T. Hallonquist and E. Suchman in *Radio Research 1942–1943* edited by P. F. Lazarsfeld and F. Stanton, pp. 265–334.

[2] Lazarsfeld, P. F., and Robinson, W. S. Some properties of the trichotomy "like, no opinion, dislike" and their psychological interpretation. *Sociometry*, 1940, 3, 151–178.

1. *Reactions to a stimulus are quantitative and in a continuous series.* Three types of reaction are obtained because of the limitation of the instructions given to the subject. But it is safe to assume that different degrees of feeling are expressed at different times within the same category of reaction. This is borne out by the comments made when the subjects talk about a program to which they have reacted.

2. *Reactions are normally distributed.* Since intensity of reactions is not measured, this assumption cannot be tested directly. However, the normal distribution remains a useful assumption, especially since it is unlikely that the distribution of reactions to a radio program ever assumes the shape of a J-curve.

3. *A point of true neutrality exists within the indifference range.* This assumption is really a specific aspect of the first two assumptions. Since it is assumed that there is a gradient of reactions ranging from extreme dislike to extreme like, there must be some point between these two extremes of true indifference or neutrality. The most likely point of neutrality seems to lie somewhere in the middle of the indifference range. Lazarsfeld and Robinson present some empirical evidence to support this assumption.[3]

The S-score itself is the distance on the base line between the neutrality point (the middle of the indifference range) to the average of the distribution, expressed in sigma units. The method of its computation is described in the previously cited article by Lazarsfeld and Robinson.

There are two apparent advantages of the S-score over the LD-score: (1) Standard deviation units are definite statistical terms with known meaning, whereas LD units are difficult to interpret. (2) Standard deviation units are equal, whereas LD units are not.

The Method

In order to study the relative merits of the S- and LD-scores for entire programs, the percentages of like, indifference and dislike were obtained for 53 different radio programs, representing a variety of program types. The number of subjects listening to a program varied from 49 to 116, with an average of 67.

The S- and LD-scores were compared for the total ratings of the 53 programs. They were also compared for program parts within three programs: one which was highly liked, one highly disliked, and one to which most reactions were indifference.

---

[3] The assumption of a neutrality point is not necessarily required when it is noted that the S-score can also be viewed as the distance of dislike in sigma units subtracted from the distance of like in sigma units. This method of calculation will yield a score exactly twice the size of the S-score.

The S- and LD-scores and the percentages of like, indifference and dislike for each of the three programs, and for the average of the 53 programs, are presented in Table 1.

Table 1

Reactions to Radio Programs

Radio Programs	Scores		Percentages			No. of Subjects
	S	LD	Like	Indif.	Dislike	
Liked	1.00	.62	69	24	7	70
Indifferent	.50	.28	37	54	9	116
Disliked	−.03	.02	25	48	27	50
Average of 53	.68	.40	40	43	9	3550

It can be seen from this table that "liked," "disliked" and "indifferent" programs are relative terms. Generally more reactions tended toward like than toward dislike, which is not unexpected since the programs are designed for entertainment. However, analyses of oral interviews of listeners to these programs confirm the designations obtained statistically.

Results

A. Relation between the S- and LD-scores. Table 2 presents correlations between the S- and LD-scores for total ratings in 53 programs and for ratings within three programs.[4]

Table 2

Correlations between S- and LD-Scores

	Correlations
Total Scores of 53 Programs	.97
Program Parts	
Liked Program	.93
Indifferent Program	.79
Disliked Program	.99

With the exception of the indifferent program, it is apparent that the S-score and the LD-score are so highly correlated with each other that the use of either score will result in approximately the same rankings of programs or of program parts. The high correlations further suggest that the data would be interpreted in much the same way whether the S- or the LD-scores were used.

---

[4] In all cases, correlations for the total ratings of the 53 programs are Pearson product-moment, and for program parts are rank-difference.

Profile charts (not shown in this report) of S- and LD-scores for program parts within the three programs show graphically the high relationship between the two scores. However, it is interesting that at the beginning and end of programs, and during applause and transition passages, where the percentage of like and dislike usually decreases, the LD-score tends to come closer to the base line than does the S-score. From this one may conclude that the LD-score reflects increased "indifference" more readily than the S-score.

B. Relations between the Two Scores and Percentage Like, Dislike and Indifference. Table 3 presents the correlations between the S- and LD-scores and the percentage of like (L), indifference (I) and dislike (D) for the total scores of 53 programs and for parts within three programs.

Table 3

Correlations between Two Scores and Like, Indifference and Dislike

	Total Scores	Program Parts		
	53 Programs	Liked Program	Indifferent Program	Disliked Program
S-L	.87	.86	.42	.68
S-I	−.62	−.75	−.09	−.32
S-D	−.80	−.27	−.64	−.80
LD-L	.96	.97	.89	.74
LD-I	−.80	−.01	−.56	−.37
LD D	−.69	.01	−.17	−.79
Minimum Significant Correlation	.33	.39	.42	.45

From this table it is apparent that the S- and LD-scores correlate very much in the same way with the percentages of like, indifference and dislike. However, the three reactions have a more equal weight in the S- than in the LD-score. This seems to be the case since the correlations for total scores and for program parts between the three reactions and the S-scores are more equal in size than for the LD-score.

Another significant difference between the two scores is that the S-score seems to give greater weight to dislike than does the LD-score. The LD-score gives the greatest weight to like, the predominant reaction.

C. The Reactions of "Indifferent" Subjects. A further comparison of the two combined scores can be obtained by examining the "indifference" reactions. One approach to this problem was to examine the reactions of the most indifferent subjects in various programs. This analysis will yield some information concerning sustained indifference at least. Those subjects who pressed either red or green buttons less than one-half of the program time were arbitrarily classified as "indifferent" subjects. For eleven different programs, the indifferent group averaged 37 per cent of all subjects.

The general attitude of the subjects to a program was ascertained by their answer to a standard question:

> In order to get you to listen to future broadcasts in this series, should the programs be: very much like this one, improved a bit, or improved a good deal?

Those who checked "very much like this one" have been designated Satisfied listeners; those who checked "improved a bit," Conditional listeners; and those who checked "improved a good deal," Dissatisfied listeners. This question has been found to correlate very highly with analyzer reactions and with the tenor of the comments made by listeners in group interviews after each program.

In eleven different programs, it was found that 27 per cent of the Satisfied listeners, 40 per cent of the Conditional listeners, and 55 per cent of the Dissatisfied listeners were "indifferent." This indicates that the "indifferent" listeners tend to be more dissatisfied with a radio program than the "non-indifferent" listeners.

One cannot conclude from these findings that "indifference" is always a negative reaction. "Indifference" may express an intermediate state between like and dislike. It may express anticipation or relaxation, as at the beginning of a program or in transition passages. However, it is clear that some so-called "indifference" reactions are really negative. This seems to be true of much of the sustained indifference. The S-score, therefore, seems to be superior to the LD-score, for radio programs at least, since it gives more weight to dislike and in that way, seems to take into account that part of indifference which is really negative.

Summary and Conclusions

The overwhelming conclusion from these data is that the S- and LD-scores yield virtually the same results and compel virtually the same interpretations of listeners' responses to radio programs. This conclusion supports the earlier investigation of Likert[5] in which he demonstrated that sigma scaling of attitude questions are no more reliable or discriminating than a scoring of 1, 2, 3, 4 and 5 for the five alternate answers. Despite this conclusion, there are certain other considerations which militate in favor of the S-score:

---

[5] Likert, R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932, No. 140.

1. *The S-score reflects all three reactions more equitably than the LD-score.* The LD-score gives the greatest weight to the predominant reaction, which in the case of radio programs is like. The S-score gives more even weight to all reactions, which in the case of reactions to radio programs, gives more weight to dislike. Such additional weight seems to be justified since it is highly likely that in a situation which is positively toned, any degree of dislike may have more significance than is indicated by a percentage. In addition, since it has been shown that some of the "indifference" reactions are really negative, dislike should be given more weight.

2. *The trend of the S-score is maintained in periods of "indifference."* The S-score does not drop as much as the LD-score at the opening and close of the program and during transition passages. The trend of the S-score seems to be more justified than the trend of the LD-score at such points because these are not really periods of indifference. Analysis of listeners' attitudes during these periods reveals that the listener is not indifferent; he is waiting for something to happen. During transition passages and at the end of the program, the listener's attitude is one of relaxation rather than indifference.

3. *The S-score units are equal whereas the LD-score units are not.* Equality of units permits direct comparison of reactions to programs and to program parts.

Received April 18, 1944.

Book Reviews

Luckiesh, M. *Light, vision and seeing.* New York: D. Van Nostrand Co., 1944. Pp. 323. $4.50.

In this treatise Dr. Luckiesh gives a simplified presentation of the relationships of light, brightness, vision, lighting and seeing. An attempt has been made to combine fundamental facts with practical discussions. The material is aimed to be helpful to those interested in better seeing conditions and their effects upon human efficiency and welfare. The book is dedicated to "Better Light—Better Sight," a slogan employed by the lighting industry.

This book will make a strong appeal to the uncritical reader or to the uninformed reader. It is of considerable importance, therefore, to examine the material in some detail. First, however, let us note some of the contributions which should receive unequivocal approval: (1) The material is clearly presented in relatively simple language. (2) The author makes a strong case for the maintenance of hygienic conditions for seeing. (3) Desirable emphasis is placed upon the relation of brightness and brightness contrast to visibility and to ease of seeing. (4) An important place is rightly given to the interdependence of the four fundamental factors (size, contrast, brightness, and time) that determine the visibility of objects. (5) The measurement of visual acuity is adequately handled. (6) One of the best sections deals with alternation of brightnesses and glare in the visual field in relation to efficiency and ease of seeing. (7) The section on light and color is practical and is effectively done.

Analysis of the discussions suggests the following criticisms to the reviewer: (1) The author consistently ignores the fact that the eyes readily adapt to easy and effective seeing over a wide range of illumination intensities. Emphasis is given only to the adaptation for effective vision at the higher levels. (2) The dismissing of findings conflicting with the author's views by means of ridicule rather than on the basis of sound criticism is both ineffective and a sign of weakness. Thus we find employed such terms as "sheer nonsense," "the valor of ignorance," "gross ignorance," "ridiculous," and "meaningless" to express the author's reaction to the contributions of others. (3) Several decades of work in the field of vision is no criterion of infallibility. Furthermore, the recurring phrases "it is axiomatic" or these "facts are axiomatic" become unconvincing, since in certain instances the critical reader will recognize that they are neither facts nor axiomatic. (4) Relatively high intensities are required for adequate seeing where discrimination of details involves low brightness contrast. It does not follow, however, that the same high intensities are necessary in the large majority of everyday visual situations. This distinction is not made clear. (5) In considering the fixational pause of the eyes, questionable data are cited although adequate data are available in the literature. (6) In citing eye-movement time for reading (page 128), the record is decidedly atypical or the computations are wrong. Duration of the back sweep and other interfixation movements are far too long. (7) It is stated that 11 point type is far above the average typography commonly encountered. This is misleading since actual figures show that journals and books are typically printed in 10, 11 or 12 point type. (8) The author's attitude toward rate of reading as an indicator of ease of seeing is highly inconsistent. Thus: (*a*) "It is axiomatic that if very poor seeing conditions are improved . . . quantity of useful work done should improve." (*b*) "As a criterion of optimum levels of illumination . . . rate of performance of a visual task is inadequate." (*c*) Nevertheless speed of reading is employed to measure influence of brightness contrast between print and background. (*d*) "*At best, speed of reading is an insensitive criterion*" of ease of reading. (*e*) But the author employs eye-movement measures (perception time, fixation frequency, pause duration, regression frequency) in reading to show the effect of variation in level of illumination. He concludes that the effects of fatigue are obviously greater for the low level of illumination as is evidenced by the change in eye-movement measures. Apparently the author does not realize that eye-movement measures are merely measures of speed of reading. It seems that rate of perception (reading) is accepted as a criterion of ease of seeing only in situations where the results support his views. (9) To measure ease of seeing the author usually obtains results for one versus 100 foot-candles, or one versus 10 versus 100 foot-candles. It is obvious that, in any visual situation where discrimination is concerned, one may expect improvement in visual efficiency in going from one to 100 foot-candles, or even from one to 10. And where discrimination is severe improvement is probable from 10 to 100 foot-candles. Luckiesh himself suggests that, to obtain a significant improvement in seeing, one should double the foot-candle level. He also states that one is generally more concerned with the practical optimum (level of illumination) than with the absolute maximum for many ordinary tasks. It is pertinent to ask, therefore, why he does not employ in his studies 1, 2, 4, 8, 16, 32, 64 and 128 foot-candles rather than only 1 versus 100, or 1 versus 10 versus 100. Responses at various levels from 10 to 100 foot-candles have not been investigated. Possibly the rate of change in efficiency is extremely slow from 15 or 20 to 100 foot-candles. In the majority of visual situations which involve details considerably above the visual threshold in size, we are interested in knowing the level of illumination above which no practical gains in efficiency occur. This cannot be revealed by Luckiesh's data. (10) The use of heart rate, the blink technique and visibility measurements as criteria of ease of seeing have been criticized in another paper.[1] An additional comment may be made. The reliability of blink scores for a five minute period of reading is low (+.49). Also one may question the stability of some differences obtained. For example, Memphis medium type is called easier to read than Memphis bold because 7 per cent more blinks occur with the bold in five minutes of reading. About 25 blinks occur in five minutes. Thus the bold would be read with 1.6 more blinks per period. With only 18 to 40 subjects, is this a stable difference? (11) The author delights in setting up straw men so that they may be knocked down. Thus he implies that someone has stated that 10 foot-candles is enough for any kind of reading—which has not been done. Then he points out the need for relatively high intensities needed for people with eye disabilities and for reading very illegible type—to which no one objects.

Dr. Luckiesh has written an interesting and important book. The treatment of several topics may be considered excellent. Much of value may be gained from the rest of the material if read critically.

Miles A. Tinker

University of Minnesota

---

[1] Tinker, M. A. A reply to Dr. Luckiesh. *J. appl. Psychol.*, 1943, 27, 469–472.

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Outlook for the serviceman: A discussion of the education, re-employment, and rehabilitation of veterans. Colonel John N. Andrews. Institute on Postwar Reconstruction, New York University, Washington Square, New York 3, N. Y., 1944. Pp. 184. $.30.

The college and teacher education. Armstrong, Hollis, and Davis. Request copies from Helen Seaton, American Council on Education, 744 Jackson Place, Washington 6, D. C., 1944. Pp. 311. $2.50.

Your problem: Can it be solved? D. J. Bradley. New York: Macmillan Co., 1945. Pp. 213. $2.00.

Counseling in personnel work. A bibliography: 1940–1944. Compiled by Paul S. Burnham. Public Administration Service, 1313 East 60th Street, Chicago 37, Ill. $1.00.

Final report on the library film forums project, 1941–43. Glen Burch, Chairman. American Library Association, 520 N. Michigan Ave., Chicago 11, Ill. Pp. 41. $.50.

Employée counseling: A new viewpoint in industrial counseling. Nathaniel Cantor. New York: McGraw-Hill Book Co., 1945. Pp. viii + 167. $2.00.

Your personality. Virginia Case. New York: Macmillan Co., 1944. Pp. 277. $3.00.

Pastoral work and personal counseling. Russell Dicks. New York: Macmillan Co., 1944. Pp. 230. $2.00.

The Brush Foundation study of child growth and development: psychometric tests. Elizabeth Ebert and Katherine Simmons. Washington: Society for Research in Child Development, National Research Council, 1943. Pp. xiv + 113. (Monographs of the Society for Research in Child Development. Vol. VIII, No. 2.)

Reading difficulty and personality organization. Edith Gann. New York: King's Crown Press, 1945. Pp. xii + 152. $2.00.

Marriage and family counseling. Sidney E. Goldstein. New York 18: McGraw-Hill Book Co., 1945. Pp. 450. $3.50.

Making and using industrial service ratings. George D. Halsey. New York: Harper & Bros., 1944. Pp. 149. $2.50.

Large Scale Rorschach techniques. A manual for the group Rorschach and multiple choice test. M. R. Harrower-Erickson and M. E. Steiner. Springfield, Ill.: Charles C. Thomas, 1944. Pp. xii + 420. $8.50.

Reality practice as educational method. Hendry, Lippitt, and Zander. New York 17: Beacon House, 1944. Psychodrama Monographs, No. 9. Pp. 36. $1.50.

Occupational therapy in the treatment of the tuberculosis patient. Holland Hudson and Marjorie Fish. Livingston, N. Y.: Livingston Press, 1944. Pp. 317. $3.00.

Effects of music on factory production. W. A. Kerr. Stanford University: Stanford University Press, 1945. Pp. 40. $1.00. Applied Psychology Monograph No. 5.

Managing your mind. S. H. Kraines and E. S. Thetford. New York: Macmillan Co., 1945. Pp. 374. $2.75.

The technique of building personal leadership. D. A. Laird. New York 18: McGraw-Hill Book Co., 1944. Pp. 239. $2.00.

The science of man in the world crisis. Ralph Linton et al. New York: Columbia University Press, 1945. Pp. xvi + 520. $4.00.

A handbook for old age counsellors. L. J. Martin. San Francisco: Geertz Printing Co., 1944. Pp. 84.

Color vision. S. D. Melville. Reprinted from four issues of *The Optometric Weekly*, 1944. $.50. Obtain reprints from Reading Clinic Sec'y, Room 8, Burrowes Educ. Bldg., Penn. State College; State College, Pa.

Soldier to Civilian. G. K. Pratt. New York 18: McGraw-Hill Book Co., 1944. Pp. 233. $2.50.

The scientific selection of salesmen. J. L. Rosenstein. New York: McGraw-Hill Book Co., 1944. Pp. 259. $3.00.

Freud: Master and friend. Hanns Sachs. Cambridge: Harvard Univ. Press, 1944. Pp. 195. $2.50.

On measurement of motor skills. E. M. Schroeder. New York: King's Crown Press, 1945. Pp. 256. $2.25.

Developing a student guidance program in an instructional department. Scott, Morgan, and Lehman. Columbus 10: Ohio State Univ. Press, 1945. Pp. 65. $.50.

Elementary educational psychology. C. E. Skinner, editor. New York 11: Prentice-Hall, Inc., 1944. Pp. 448. $3.75.

The handbook of industrial psychology. May Smith. Philosophical Library, 15 E. 40th St., New York, N. Y. Pp. 304. $5.00.

Role analysis and audience structure. Zerka Toeman. New York 17: Beacon House, 1944. Psychodrama Monographs, No. 12. Pp. 19. $1.25.

Personnel relations: Their applications in a democracy. B. J. E. Walters. New York: Ronald Press Co., 1945. Pp. 547. $4.50.

First course in psychology. R. S. Woodworth and M. R. Sheehan. New York: Henry Holt, 1944. Pp. x + 445.

Normal lives for the disabled. Edna Yost and Lillian M. Gilbreth. New York: Macmillan Co., 1944. Pp. 298. $2.50.

Guide to the evaluation of educational experiences in the armed services. Compiled for the American Council on Education under the direction of G. P. Tuttle. $2.00 a set. Mail orders to 363 Administration Bldg., Urbana, Illinois.

Music in industry: A manual on music for work and for recreation in business and industry. Industrial Recreation Association, Chicago, 1944. Pp. 64.

Personnel records. Industrial Welfare Society, Inc., 14 Hobart Place, London, S.W. 1, England, 1944. Pp. 24. 2s.

The Van Allyn job placement technique. National Institute of Vocational Research, 305 W. 8th St., Los Angeles 14, California. $10.00.

Journal of Applied Psychology

Vol. 29, No. 4 August, 1945

Intelligence and Adjustment Measurements in the Selection of Radio Tube Mounters

George Forlano [*] **and Forrest H. Kirkpatrick**
Radio Corporation of America *Bethany College*

The general problem with which this study was concerned is the increase of worker efficiency in radio tube mounting jobs. Specifically the problem reduced to: will intelligence and adjustment tests be effective in bringing about this desired increase in efficiency? There have been many claims as to the effectiveness of adjustment tests in particular in the selection of more desirable workers, but there have been too few quantitative presentations of proof. Vague opinions of personnel officers that the use of tests brings superior employees are too often accepted without evidence.

The subjects employed in this experiment were twenty female tube mounters in one department of the Indianapolis Plant of the Radio Corporation of America. Tube mounters assemble the very small tube elements to the bases of radio tubes. It is very close work requiring considerable finger and hand dexterity. Several mounters, each performing distinct operations, work together on each tube. The subjects were new employees who had taken the required placement tests, and had been assigned as tube mounters. New employees were used in the experiment because it was desirable to keep all regular workers on the production line.

The tests used for the purposes of this experiment were the (1) Otis Self-Administering Test of Mental Ability Form B, (2) the Bell Adjustment Inventory, adult form, and (3) the Washburne Social Adjustment Inventory. Of the latter two tests only the scores which are designed to measure social adjustment were used. This included the "social" score of the Bell test and the "alienation" score of the Washburne test. These were administered before the employees started work. The subjects also took the regular aptitude battery (a vision test, a two-hand coordination test, and manual dexterity tests) with which we shall not concern ourselves in this report.

---

[*] Now in the U. S. Army.

The criterion used for this experiment was ratings by the supervisor in charge of the group. Each worker was rated for the extent to which she took hold of the job and performed efficiently. These ratings were given on the basis of one month's observation of the worker at her task. The supervisor's ratings were in the following terms: G, indicating a good employee; and F, indicating a fair employee.

For purposes of examination of the data, test score distributions were divided into three groups and scores within the groups assigned the values: above average, average, or below average. Table 1 gives the groupings

Table 1

Distribution of Test Score Values to Various Groupings

Test	Grouping		
	Above Average	Average	Below Average
Otis Mental Ability (I.Q.)	115 and over	95 to 114	94 and below
Bell Social Scores	5 to 8	9 to 19	20 to 24
Washburne Alienation Scores	0 to 5	6 to 15	16 and above

Table 2

Showing Mounters' Tests and Related Data

Tube Mounter	Supervisor's Ratings	Otis I.Q.'s	Personality Rating	Composite Intelligence and Personality
1.	G	125	+av.	12
2.	F	122	−av.	6
3.	G	122	av.	9
4.	G	122	+av.	12
5.	F	118	−av.	6
6.	G	118	av.	9
7.	F	118	av.	9
8.	F	114	−av.	3
9.	G	111	+av.	9
10.	F	101	−av.	3
11.	G	98	av.	6
12.	G	97	+av.	9
13.	F	95	av.	6
14.	F	91	av.	3
15.	F	80	−av.	0
16.	F	76	−av.	0
17.	F	70	av.	3
18.	F	72	−av.	0
19.	G	69	av.	3
20.	F	95	av.	3

used. These divisions were based on distribution of scores into upper, middle, and lower thirds.

The two personality test scores were then combined in such a way that if a worker made above average on both tests she was rated "+ average" for personality; if she received above average on one test and below on the other, or if she was average on both, she was rated "average" in personality. Likewise, if the worker made below average on both tests she was rated "−average" in personality. These are the figures shown in Table 2 under the heading "Personality Rating."

Table 2 also shows a column headed "Composite Intelligence and Personality." These figures were obtained by weighting an above average personality rating or Otis score as 6, an average rating or score as 3, and a below average rating or score as 0, and then adding the two weights for each person. Hence, the composite could range from 0 (below average Otis score + below average personality rating) to 12 (above average Otis score + above average personality rating). Table 2, then, contains the basic figures from which experimental results were studied.

Intelligence and Job Success

Table 3 shows the relationship of intelligence test scores to the ratings of job success. For employees with average and above average scores there were exactly the same number rated "good" by supervisors as there were rated "fair." However, 6 out of 7 who were below average in intelligence had been rated as "fair" by the supervisor. It would seem from these results that low intelligence would go with poor job success while above average intelligence does not give advantage over average. This might have been expected since, although it takes a certain amount of intelligence to learn the job, there is not much to tax mental capacity after it is learned.

Table 3

Distributions of Intelligence Test Scores for "Good" and "Fair" Employees

Test Scores	Rated "Good"	Rated "Fair"
Above Average	3	3
Average	3	3
Below Average	1	6

Social Adjustment and Job Success

Table 4 summarizes the relationship between personality test scores for social adjustment and ratings of job success. The scale indicating "social adjustment" here is the combination of Bell and Washburne scores given under "Personality Rating" in Table 2. These results are rather striking in that all of the people scoring above average on the tests also were rated "good" by their supervisor, while all of the people scoring below average on the tests were rated "fair" by the supervisors. Whether this indicates that a socially adjusted personality is an important factor in doing good work, or that the supervisor was more influenced by the girls' "personalities" than their performance, is uncertain. Certainly the tests were measuring something which was related to the supervisor's notion of good and fair employee risks.

Table 4

Distributions of Personality Ratings for "Good" and "Fair" Employees

Test Scores	Rated "Good"	Rated "Fair"
Above Average	4	0
Average	4	5
Below Average	0	7

Job Success and the Composite Score

When the intelligence and personality scores are combined in the manner indicated in Table 1, distinction between "good" and "fair" employees is apparent. Table 5 shows that by eliminating all the employees who had a composite score of 3 or less, we would have left 88% of the "good" employees and only 39% of the "fair" employees.

Table 5

Composite Intelligence and Personality Score Distributions for "Good" and "Fair" Workers

Composite Score	Rated "Good"	Rated "Fair"
12	2	0
9	4	1
6	1	4
3	1	5
0	0	3

Moreover, by eliminating employees whose composite scores are 6 or less, all but one, or 92%, of "fair" employees would be eliminated. At the same time only 2, or 25%, of "good" employees would be lost. It would seem from these figures that, in times of a favorable labor market, these tests would be very effective in selecting workers able to meet with the approval of supervisors.

Summary

This study was concerned with the effectiveness of certain personality and intelligence measures in predicting the job success of 20 tube mounters. Relationships were shown between supervisory ratings of job success and scores on the Otis Self-Administering Test of Mental Ability, Form B, the social scale of the Bell Adjustment Inventory, and the alienation scale of the Washburne Social Adjustment Inventory. The following conclusions were drawn:

(1) Low intelligence scores tended to indicate the poorer workers. Average or above average scores did not discriminate between "good" and "fair" workers.

(2) All workers scoring below average in social adjustment had been rated only "fair" by the supervisor while all workers scoring above average in social adjustment had been rated "good" by the supervisor. The supervisor had no previous knowledge of test scores.

(3) A composite of intelligence and personality scores was shown to be effective in predicting the subsequent success of new tube mounters.

Received July 28, 1944.

Single-Item Tests for Psychometric Screening *

Lieut. H. M. Hildreth, USNR

A series of 10 Single-Item Tests are presented in this article. These tests, designed for psychometric screening, were developed at a Naval training station in response to the need for rapid methods of examining the intelligence of large numbers of recruits. They take from 1 to 2 minutes per man to administer, and do not require optimal testing conditions. Standardized on 1,500 cases they have also been tested in actual practice and found to be practicable. In a 3-month trial period, during which the tests were used as a basis for accepting recruits for service, the error involved was found to be less than 1/10 of 1 per cent.

A New Approach to the Problem of Screening

Single-Item Tests are the result of a new approach to the screening problem of increasing speed of testing without losing accuracy. The usual approach has consisted of abbreviating, and devising short forms of, standard mental tests. These short forms bring testing time down from an hour to around 10 or 15 minutes, without appreciable loss of accuracy; but this is about the limit to which standard-type tests can be skeletonized without sacrificing reliability.

The reason for the limitation of this approach lies in the psychometric requirements imposed by the purpose of mental measurement. In the past this purpose has always been to determine the true or *maximum* mental ability of the individual. Standard tests, including those recently streamlined for screening, have always been designed and constructed to measure this maximum.

In psychometric screening the individual's maximum is not a matter of concern. The purpose in screening is to make sure that a recruit's mental ability is not below a given *minimum* and that he is thereby acceptable for naval or military service. Tests for acceptance need measure only this minimum; it is not necessary to measure maximum ability.

Recognition of this basic difference in purpose makes it possible to dispense with 3 traditional psychometric requirements. In tests for maximum ability: (1) many items are ordinarily included in the test so that the individual will have ample opportunity to demonstrate his full ability; (2) failure is accorded as much significance as success, since the failure level of an individual must be determined if his maximum performance is to be known; and (3) scoring of failure requires in turn that maximum-ability tests be administered under optimal testing conditions. Unlike success, which demonstrates the presence of ability, failure does not necessarily indicate lack of ability. It may be due to distraction, confusion or anxiety, and until these and similar extraneous causes are eliminated, failure cannot be taken to mean "lack of ability."

---

* The opinions or assertions contained in this article are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

In testing for minimal mental ability none of these requirements need be considered. One success is sufficient to establish a minimum. A recruit who is successful on a single item demonstrates he has at least the degree of mental ability called for by that item. He may well have more, but it is certain he does not have less. If the measuring power of the particular item has been determined, then a minimal mental ability can be ascribed to the man at once and without further testing.[1]

Ordinarily, however, only the average difficulty of test items is known. The minimal mental ability which they represent is not known. To obtain this information, and make possible the use of single items as screen tests, the special scoring technique described below was devised.

Scoring Technique

The technique used in scoring items for screening consists of tabulating for a normal group the mental ages of all those who are successful on a given item, and then determining the point at which this distribution becomes asymptotic to the mental-age base line. Those items found to represent a minimal mental ability of the desired degree are suitable as screen tests.

The desired minimum in the present case was 11 years, or 132 months, mental age, the minimal mental ability for Naval service originally set by official directives. Thirty items from well-known tests (1, 2, 3) were selected for scoring. In conjunction with a Stanford-Binet examination these 30 items were administered under optimal conditions to 1,500 men of military age. Care was taken to assure a normal distribution in respect to mental age, particularly in the lower brackets.

Each of the 30 experimental items was then scored for minimal mental ability. (1) A frequency distribution was made of the mental ages of all those who passed the given item. Failures were disregarded. (2) The point at which the distribution approached zero was located. For practical purposes this point was set at the first percentile, with the one-half percentile point also being computed. (3) The mental ages corresponding to these points were determined. If the mental age at the first percentile was 132 months or greater, the item was considered suitable as a Single-Item Test for screening.

---

[1] The foregoing applies to free-response items such as those used in the S-I Tests. Success cannot be considered prima facie evidence of ability in multiple-choice or true-false items where an appreciable element of chance is present. Coaching, of course, invalidates any mental test.

In the "Squares" problem below, for example, all individuals in the standardization group passing the item were separated out and their mental ages were tabulated in a frequency distribution. The first percentile of this distribution was computed and found to be 132 months. Since this value was not below the desired minimum, the item was considered suitable as S-I Test. The chances are less than 1 in 100 that a recruit who passes this test during screening will have a mental age below 132 months.

Of the 30 experimental items scored in this manner 10 were found to guarantee mental ability above the Navy minimum. These 10, designated as Single-Item Tests, are described below.[2]

Description of Single-Item Tests

In the column headed $P = .990$ is listed (in months) the mental age value at the first percentile. The probability is .99 that an individual passing the S-I Test will have a mental age at or above that listed in the table. For column $P = .995$ the chances are 199 in 200 that the individual will not be below the mental age listed there. These probabilities are in terms of a normally distributed male group of military age, as found at an induction center. This is the group customarily dealt with in screening.[3]

Scoring of the Kent and Stanford-Binet items is in most instances identical with that originally described for these items.

1. Question:	"How much is:	$P = .990$	$P = .995$
7 times 7	10 times 10	132	128
8 times 8	11 times 11		
9 times 9	12 times 12		

Scoring: Pass if subject makes no more than two errors.

---

[2] It will be noted that some of the values in this table are below 132 months, but not below 126. It has been found in practice that 126 is a safe value to use because of the conditions under which screening is usually done, and because Stanford-Binet scores on subnormals tend to run somewhat lower than the Wechsler-Bellevue (4) which is frequently used for determining mental age. Scoring of Wechsler-Bellevue items for screening is now under way.

[3] When testing is limited to doubtful cases all of whom are suspected of mental deficiency success may be conservatively interpreted in terms of this group. The probability is less than .05 that an individual whose mental age is below 11 years will pass any given S-I Test. The null-type hypothesis that the individual is defective is then disproved by his success.

2. Question: "What does the word PRICELESS mean? 133 126

Scoring: Full credit for such responses as "very valuable," "you can't buy it for any amount of money," "it is worth a lot." If the subject says "It has no price," ask what he means. Make sure he really knows the meaning of the word.

3. Question: "If your shadow points toward the northeast, in which direction is the sun?" 134 128

Scoring: Credit if response is "Southwest."

4. Question: "What does the word TOLERATE mean?" 142 139

Scoring: Full credit for such responses as "it means to put up with things," "it means you stand for what people do without getting mad," "it means to be lenient," "to endure." Do not credit such responses as "I tolerate you" unless the subject can follow with a clear explanation. Do not credit such responses as "it's a sickness—the cholery."

5. Question: "What is the next number?"
   1 2 4 8 16 — 139 137

These numbers are presented in written form.

If subject hesitates, say "These numbers go up in a certain order; what would be the next one?"

Scoring: Credit only if the subject says "32" and is able to explain that the numbers are being doubled.

6. Question: "Why does the moon look bigger than the stars?"
   "What time of day is your shadow shortest?" 136 132

Scoring: Full credit if both questions are answered correctly.

For the first question credit such responses as "the moon is closer (nearer) to the earth," "the stars are farther away." No credit for such answers as "the moon is bigger"; give assurance that the moon is really smaller, and repeat the question.

For the second question, credit "noon" or any time between 11:00 and 1:00. No credit for "night" or "4 o'clock."

7. Question: "What is the opposite of WINTER?" "What is the opposite of WAR?" "What is the opposite of BROAD?" "What is the opposite of ALIKE?" "What is the opposite of DEEP?" 135 129

Scoring: Full credit if subject can give the opposite of four of these words.

8. Question: "Make a good sentence out of these words." 129 126
   FOR THE STARTED AN WE COUNTRY EARLY AT HOUR

These are presented to the subject in written form.

Scoring: Full credit for "we started for the country at an early hour," "at an early hour we started for the the country," "We started at an early hour for the country."

9. Question: "What does the word MARS mean?" "What does the word HYSTERICS mean?" 130 127
   "What does the word BRUNETTE mean?"

Scoring: Pass the subject if the meaning of each of the words can be given. For MARS give credit for "planet" or "the God of War." No credit for "it is up in the sky," "it is a place," "it is a country," "it is a candy bar." For HYSTERICS give credit for "it is nervous sickness," "it is like having a fit of crying," "to get excited and go all to pieces." No credit for "it is about old times" (historic). For BRUNETTE give credit for "dark complexion," "black hair," "a person who is dark." No credit for "a blonde," "it is a woman," "red hair."		
10. Question: "In what way are a KNIFE-BLADE, a PENNY, and a PIECE OF WIRE alike?"	129	126
Scoring: Full credit if response is "metal." No credit for "copper," "steel," "they are useful," or "they're all hard."		
	$P = .990$	$P = .995$

Use of S-I Tests in Screening

Each of the 10 S-I Tests above is an independent test of acceptance. Success on any one of the tests demonstrates mental ability above the minimum for Naval training.[4] No attention need be given to providing optimal testing conditions. A distracting situation does not invalidate an S-I Test; on the contrary it increases the significance of the recruit's success. When privacy is not available and the replies of one man can be overheard by those next in line, the tests can be used in rotation, each man receiving a different question.

The majority of men pass the first test given them promptly. Failure on any or all of the S-I Tests, however, does not mean that a recruit is necessarily below the required mental minimum. Confusion rather than lack of ability may cause the failure. If the recruit is unsuccessful in several tries, he should be held over for further examination.

This subsequent examination may include more elaborate phases of acceptance testing (5). It may also involve testing for rejection with the usual full-length psychological tests. These "many-item" tests, designed to measure maximum intelligence, are necessary before a decision to reject can be made, and the more hours available for thorough rejection-testing of this group the more men can be salvaged for service. Single-Item Tests, by speeding up acceptance, save time for the more careful examination of these borderline cases.

---

[4] Present data indicate that when testing is continued until 2 S-I Tests are passed, 3 months may be added to each of the lower values. E.g. if the first and second tests are passed the new minimal values would be 135 and 129 for the 2 degrees of probability. In some cases, but not all, this will establish a higher minimum than that given for either of the 2 tests.

Summary

Ten Single-Item Tests for use in psychometric screening are described. These tests, the result of a new approach to the problem of screening, require 1 to 2 minutes per man to administer. Success on any one of them denotes mental ability above the minimum required for Naval training.

S-I Tests for rapid acceptance allow more time for rejection-testing of borderline cases, many of whom can be salvaged when additional examining time is available.

Received July 5, 1944.

References

1. Terman, L. M. *The measurement of intelligence.* New York: Houghton Mifflin Co., 1916.
2. Kent, G. H. *Oral test for emergency use in clinics.* Baltimore: The Williams & Wilkins Co., 1932.
3. Brown, M. A simple method for rapid estimation of intelligence in adults. *Amer. J. Orthopsychiat.*, 1942, 12, 411–414.
4. Wechsler, D. *The measurement of adult intelligence.* Baltimore: The Williams & Wilkins Co., 1941.
5. Hildreth, H. M., Wheeler, J. A., and Williams, S. B. *A psychometric procedure for screening mental defectives.* U. S. Naval Medical Bulletin (in press).

Range of Interests *

Ralph F. Berdie

Lieut. (jg), U.S.N.R.

Everyday observation and clinical experience suggest that the well adjusted individual has a wide range of interests and that other individuals who are experiencing conflicts within themselves and with their social environments tend to have more restricted ranges of interests. Narcissism accompanies many psychological malfunctions, and inability to attend to and maintain interest in occupational and extra-vocational activities often characterizes the neurotic adult. Extreme egocentricism is normal in infants and very small children. In physically mature individuals it indicates a lack of personality development or a regression resulting from various psychological frustrations. This regression may be as complete as that we observe in the advanced case of schizophrenia, or it may be more subtle.

Experience again suggests that the ranges of interests possessed by normal persons may vary widely and that some otherwise abnormal persons may have interests that do not differentiate them from normal people. A suspicion is justified that a few abnormal persons might have unusually wide ranges of interests and that this extensity of interest may or may not be related to their abnormality.

Very little relevant, systematic information is available regarding this problem of the relationship between the range of interests and other psychological attributes. The range of interests, as measured by the number of items marked as liked or disliked on the Strong Vocational Interest Blank, was found in one study[1] to be slightly correlated with high school and college grades and morale and social adjustment scores of the Minnesota Personality Scale. These correlations, ranging from −.18 to +.23 and determined on a sample of 411 college freshmen, were statistically significant but were too small to be of any value in predicting adjustment. The items liked on the Strong test are determined to a far greater extent by the person's vocational interest profile than by his personal adjustment. That is why the items were selected. The relationship found between range of interests and adjustment, however, suggests that items selected with the purpose of predicting adjustment rather than differentiating vocational interests might have a greater prognostic value.

---

* The opinions or assertions contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

[1] Berdie, R. F. Likes, dislikes, and vocational interests. *J. appl. Psychol.*, 1943, 27, 180–189.

The problem of predicting personal adjustment is particularly important to military psychologists attempting to eliminate those misfits from the service who, if allowed to remain, will eventually break down so as to endanger themselves and their fellows and shoulder the government with rehabilitation and pension costs.

Each man entering military service passes through at least one psychiatric screening. The type of screening depends on the available staff, the time allowed, and the facilities at hand, but in each case an attempt is made to determine if the man presents an adequate military risk or not. The pressure of work requires that the psychiatric screening be quick and even in those places where time for more extensive study is available, complete personality studies are not feasible.

Mental deficiency and illiteracy can be determined with relative simplicity. The frank psychotic, the organic deteriorate and the obvious misfit can be selected with but little more difficulty. Other cases, such as anxiety neurotics, hysterical personalities and constitutional psychopathic states, are more difficult to identify.

The problem attacked in this study concerned the usefulness of the range of interests in selecting potential psychological misfits at the beginning of military training. Does the range of interests possessed by an individual bear a relationship to his military adjustment and can this range be adequately measured? Are criteria available against which we can evaluate the predictive value of such a measure and finally, is such a measure useful in the military situation?

The study is divided into two sections. The first concerns a list of interest items presented orally to recruits. Here are discussed the questions of how stable the likes are from group to group, how the range of interests is related to ability and how well it differentiates between normal and abnormal recruits.

The second section of the study concerns a list of interest items presented in printed form to recruits. Here are discussed the questions of how reliably this list measures the range of interests, what the relationship between the range of interests, age, and education is, and how well it differentiates between normal and various groups of abnormal recruits.

Orally Presented List

A group of 31 activity interest items were pre-tested on a group of marine recruits and 22 of these items were liked by over 50 per cent of the men. These 22 items were retained to form the interest scale. The items are presented in Figure 1.

Name———————————Age———Platoon No.——————

Highest school grade completed:————at———— years of age.

DIRECTIONS: Below is a list of things people do for recreation. Place a check (X) after the name of each thing you *LIKE TO DO*. After those things you dislike or have never done, leave a blank space.

Play checkers............ ————
Play pool................ ————
Play horseshoes........... ————
Dance.................... ————
Go on dates.............. ————

Box...................... ————
Play cards............... ————
Go to parties............ ————
Play basketball.......... ————
Play baseball............ ————
Play football............ ————

Fishing.................. ————
Hiking................... ————
Go to carnivals.......... ————
Go on picnics............ ————
Reading.................. ————

Go to movies............. ————
Listen to radio.......... ————
Roller skate............. ————
Watch baseball........... ————
Watch football........... ————
Go swimming.............. ————

FIG. 1. Interest list presented to marine recruits immediately after physical examination.

The items were then presented in individual interviews to 200 marine recruits selected at random at the completion of their physical examination.[2] In the interview the examiner said,

> "I would like to find out what things you like to do. I will name these things to you and you tell me if you like to do them or not."

The items were presented in the order they appear in Figure 1, the examiner saying, "Do you like to play checkers?", "Do you like to go out on dates?", etc. With the last item, if the man said he was married, he was asked, "Did you like to go out on dates before you were married?" If the recruit said he liked the activity or disliked it, the examiner recorded the response. If he said he did not know or had no contact with the activity, this was recorded as an "indifference" response. Some subjective judgments were necessary in recording the responses but any statement connotating acceptance ("I like it sometimes") was recorded as a "like."

The list of items was then presented to 68 marine recruits recommended by the psychiatric unit to the aptitude board for inaptitude discharges. These recruits were all literate and none were mental defectives.

---

[2] The representativeness of this group is evidenced by the mean General Classification Test score of 101.8, SD = 16.3 for 172 of the men for whom G.C.T. scores were available.

Their difficulties included anxiety neuroses, hysterical states, epilepsy, enuresis, constitutional psychopathic states, post-traumatic syndromes, and psychasthenia. They were a very heterogeneous group of non-psychotic psychiatric cases.

Each man had been interviewed at least twice by a psychiatrist and twice by a psychologist and psychological test scores and verified social histories were available for many. In each case the decision had been made that their problems would prevent them from making an adequate military adjustment. These decisions were all made before the presentation of the interest list.

Two indices of range of interest were obtained from the responses to the items. The first consisted only of the number of items liked. The second was based upon weights assigned to each item upon the basis of the percentage of the number of cases liking the item. A man not liking to listen to the radio, an activity liked by 98 per cent of men, is more of a deviate than one who doesn't like to dance, an activity liked by only 53 per cent of men. By weighting each response in direct proportion to the extent to which it indicated an individual was a deviate, we hoped to obtain a more differentiating score. The results showed that the scores based upon the weights gave no better differentiation than the scores consisting merely of the number of activities liked.

Results

The stability of interests from group to group is important. Will interest items rank themselves in the same order of popularity in different samples or do conditions affecting the popularity of different activities vary from group to group to such an extent that interest in these activities cannot serve as a basis for constructing a scale for general use? Are there items which have a well defined place in the recreational hierarchy of adult males in America?

The 22 items were ranked according to the proportion of the first 50 subjects who reported liking the items and then ranked according to the proportion of the second 50 subjects who liked the items. The rank order correlation was .89. The same thing was then done on the basis of the likes of the first and second one hundred subjects and again the rank order correlation was .89.

In almost any group of American men selected at random, between 90 and 100 per cent will report they like to go to the movies, about one-half will report they like to dance. As reported later, age and perhaps educational status help determine the individuals' interests, along with multitudinous other factors, but in considering groups drawn from the general population, these interests form themselves into a remarkably stable order.

General intelligence has frequently been mentioned as one of the factors determining interests. In order to see what relationship existed between the total number of likes reported by an individual and his intelligence, comparisons on the basis of the number of likes was made between groups receiving high, medium and low scores on the General Classification Test. Test scores were available for 172 of the 200 men selected at random. They were divided into three groups; 37 who had G.C.T. scores of 110 and above; 47 who had scores of 89 and below, and 88 who had scores of between 90 and 109 inclusive. The cumulative percentage distributions of the number of likes for the three groups are presented in the first three columns of Table 1.

The average person in the low ability group likes 18 plus items. The average person of the medium ability group likes 18 plus items. The

Table 1

Distributions of Number of Likes Obtained with Orally Presented List for Three Different Ability Groups, Group of Discharged Recruits and Normal Recruits

No. of Items Liked	G.C.T. = 110 and Above		G.C.T. = 89 and Below		G.C.T. = 90-109		Inaptitude Discharges		Normal Recruits	
	Freq	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %
22	1	100	1	100	5	100			7	100
21			3	98	11	95			15	96
20	6	97	5	92	16	82			30	80
19	2	81	9	81	9	64	1	100	21	74
18	9	76	6	62	11	54			35	63
17	4	51	5	40	10	41	5	98	22	46
16	7	41	6	30	11	30	2	91	30	35
15	4	22	6	26	3	17	1	88	14	20
14	3	11	2	13	5	14	3	86	10	13
13					2	8	7	82	3	8
12	1	3	1	9	2	6	9	72	5	7
11							8	50		
10			1	6	1	3	9	47	3	4
9			2	4	1	2	4	34	3	2
8					1	1	5	28	2	1
7							4	20		
6							3	15		
5							2	10		
4							2	7		
3							2	4		
2							1	1		
Total	37		47		88		68		200	

average person of the high ability group likes 17 items. Inspection of the columns reveals little or no significant difference between the number of items liked by the three groups. This is particularly evident when the figures for the ability groups are compared with the figures for the discharged and normal groups, which will be discussed later. The conclusion seems warranted that, with this particular list of items, the range of interests bears little relation to general intelligence.

The primary question about the list of items is how well does it differentiate between people who are making a good adjustment to military training and people who are not. The answer is found in the last two columns of Table 1.

The column to the left presents the accumulative percentile ratings of the 68 inaptitude cases and the column at the right the same thing for the 200 normal cases. These figures are based on simply the number of likes reported by the individuals. The range for the inaptitude cases is from 2 to 19, with fifty per cent of the cases liking 10 or fewer items. For the normal cases, the range is from 8 to 22, with fifty per cent of the cases liking 18 or fewer items. Of the inaptitude cases, 20 per cent like fewer items than any of the normals while of the normals, 37 per cent like more items than do any of the inaptitude cases. By choosing a critical point at 13 items, we would eliminate only 8 per cent of the normals and 82 per cent of the inaptitude cases. Such a small degree of overlapping is almost unprecedented when psychological instruments are used. Little question can be raised that the normal and atypical groups studied here most certainly differ on the basis of the number of likes they report.

Printed Interest List. The preceding discussion concerns data collected in a brief, standardized interview. The collection of information in personal interviews has certain disadvantages, however, when dealing with large groups. As has already been mentioned, this method in a few cases requires the interviewer to make judgments as to whether the subject's response indicates liking or not. To facilitate the collection of information and to eliminate any variability associated with the interviewer, the interest list was presented as a check list. A copy of the list and instructions as presented to the subjects is presented in Figure 1.

The mimeographed interest list was presented to 792 marine recruits, all the members of eleven consecutive platoons, immediately upon the completion of their physical examination and just prior to their interviews with the psychiatrists. Of these 792 men, 23 were selected by the psychiatrists to return for further examination because of suspected illiteracy and/or mental deficiency. Another group of 25 recruits was selected to return because of suggestions of other psychiatric or psychological defects —psychoneurosis, epilepsy, etc.

The check list was also presented to 114 recruits who had been thoroughly studied by the psychiatrists and psychologists and who were being recommended to the aptitude board by the psychiatric unit for inaptitude discharge. These were almost all of the recruits not of defective or borderline intelligence who were recommended to the board over a certain period of time. They were given the list after the final decision had been made to recommend their discharge. All of these recruits had at least two interviews with a psychiatrist and one with a psychologist. Most of them had been seen twice by a psychologist. Many had several contacts with both psychiatrists and psychologists and many had been carefully observed in a special observation unit or carefully studied in the neuropsychiatric ward of a Naval hospital. Social histories available for most of the recruits substantiated the information which led to the recommendation for their discharge. This group of discharged individuals included recruits diagnosed as having anxiety neurosis; constitutional psychopathic state, emotional instability; constitutional psychopathic state, inadequate personality; psychoneurosis, mixed; post-traumatic syndrome; epilepsey; enuresis; psychoneurosis, hysteria; constitutional psychopathic state, schizoid personality; hypochondriasis; psychoneurosis, psychasthenia; somnambulism; constitutional psychopathic inferior without psychosis; psychoneurosis, neurasthenia; cryptic nostalgia.

Results for the Printed Interest List. The range of scores obtained with the mimeographed list was greater and the distribution less skewed than with the scores collected in the interview. Therefore the calculation of a reliability coefficient was more feasible for the written list than for the other. From the total group of 792 cases, 100 cases were selected and the odd-even reliability was computed on the basis of the number of likes checked on the even numbered items and the number of likes checked on the odd numbered items. The corrected coefficient was .88. Very few psychological tests with only 22 items have reliabilities high enough to warrant their use with individuals.

Vocational interests are known to be influenced by age [3] and common sense suggests recreational interests are equally as, if not more, susceptible to the influence of increasing age. The ages of 780 of the 792 subjects were known. The mean age of this group was 24.4 years, with a standard deviation of 7.9 and a range from 17 years to 49 years. The total distribution was extremely skewed with 318 of the cases falling within the 17–18 year old group. The product-moment correlation of −.13 between age and range of interests does not offer a true estimate of this relationship because of the skewness.

---

[3] Strong, E. K., Jr. *Change of interests with age.* Stanford, Calif.: Stanford Univ. Press, 1931.

A better estimate of the influence of age upon range of interests was obtained by comparing the number of items checked as liked by the 406 men in the 17 to 22 year old group to the number liked by the 177 men in 29–49 year old group. The distributions for these two groups are presented in Table 2.

Table 2

Distributions of Number of Likes Obtained with Printed List for Different Age and Educational Groups

No. of Items Liked	17–22 Yr. Olds		29–49 Yr. Olds		17–22 Yr. Olds 8th Grade or Less		17–22 Yr. Olds H.S. Grad. or More	
	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %
22	32	100	10	100	1	100	10	100
21	21	94	5	94	4	99	5	90
20	30	88	7	92	7	94	7	84
19	22	81	7	88	4	86	9	77
18	41	75	9	84	7	81	11	68
17	28	65	6	79	4	73	8	57
16	31	58	4	75	8	68	9	48
15	30	50	6	71	8	59	4	39
14	20	43	14	70	7	50	11	35
13	22	35	8	62	8	41	4	24
12	24	30	10	57	5	32	5	20
11	20	24	15	51	4	26	3	14
10	13	19	13	43	3	21	2	11
9	17	16	9	36	3	18	2	9
8	17	12	10	31	1	14	4	7
7	15	7	9	25	4	12	1	3
6	5	4	7	20	4	8	1	2
5	5	2	7	16	3	4		
4	3	1	12	12			1	1
3	1	1	5	5				
2								
1			4	2				
Total	406		177		85		97	

In the younger group, 50 per cent of the men liked 15 or more of the activities. In the older group, only 30 per cent liked 15 or more items. *In the younger group, 12 per cent liked only 8 or fewer items, while in the* older group, 31 per cent liked no more than this many activities. Inspection of the table reveals a substantial decrease in the number of activities liked as one moves from the younger to the older group. Table 3 presents the per cent of people in each of these two age groups and in the groups divided upon the basis of education who check each item. In almost every case, the difference agrees with expectation.

Age brings a decreasing interest in dates. The younger men like to play horseshoes more than the older but this difference is influenced by a greater preference of the younger, less well educated group. Age apparently bears little influence on interest in dancing or on interest in playing cards. Younger men are more interested in parties, basketball, football and hiking. Older men are slightly more interested in watching baseball and in reading. Inspection of the table indicates that changes in interest accompanying age are not independent of the educational status of the individual.

The above results show that when an individual's range of interests is to be used as a personality index, his age must be considered. A man of forty years cannot be compared legitimately with a youth of eighteen.

Table 3

Percentage of Various Groups Checking that They Liked Each Item

Item	17–22 Yrs. 8th Gr. or Less	17–22 Yrs. More than 8th Gr.	23–40 Yrs. 8th Gr. or Less	20–40 Yrs. More than 8th Gr.	17–40 Yrs. 8th Gr. or Less	17–40 Yrs. More than 8th Gr.
Play checkers	71	62	50	62	61	63
Play pool	67	70	55	66	63	66
Play horseshoes	72	57	57	59	67	61
Dance	40	58	41	59	41	59
Go on dates	83	83	27	39	53	68
Box	46	48	22	32	36	42
Play cards	58	60	59	60	59	61
Go to parties	62	72	37	44	49	62
Play basketball	55	70	18	44	34	61
Play baseball	59	77	55	64	63	73
Play football	55	77	22	44	39	60
Fishing	76	76	71	78	76	77
Hiking	57	59	40	48	48	55
Go to carnivals	58	68	44	41	53	58
Go on picnics	54	62	52	49	53	58
Reading	63	71	63	82	63	75
Go to movies	90	95	77	88	86	94
Listen to radio	83	84	75	86	81	86
Roller skate	48	55	38	33	41	48
Watch baseball	70	71	75	79	75	73
Watch football	70	81	54	72	63	80
Go swimming	79	84	69	77	75	83
No. of cases	85	289	89	115	229	483

Table 3—*Continued*

Item	17–22 Yrs.	29–49 Yrs.	Inaptitude Discharge Cases	Normal Matched Controls
Play checkers	64	57	46	49
Play pool	60	55	24	62
Play horseshoes	61	58	32	67
Dance	54	51	27	47
Go on dates	83	34	27	50
Box	47	28	11	31
Play cards	59	60	37	61
Go to parties	69	41	24	48
Play basketball	67	32	14	37
Play baseball	73	60	36	65
Play football	71	34	26	39
Fishing	76	75	62	78
Hiking	59	45	17	37
Go to carnivals	65	42	27	44
Go on picnics	66	50	38	55
Reading	69	74	44	71
Go to movies	94	83	59	89
Listen to radio	84	81	79	81
Roller skate	53	35	16	46
Watch baseball	70	77	46	72
Watch football	78	64	30	66
Go swimming	82	74	34	83
No. of cases	374*	204*	144	144

* These total numbers of cases do not correspond with the totals presented in Table 2 as some cases for whom complete data were not available were discarded and additional cases were obtained to increase the size of the older group.

Separate normative tables must be used in order to provide a relevant frame of reference.

Range of interests might also vary with the educational status of the individual. In order to test this hypothesis a group of high educational attainment was compared with a group of lower educational attainment, holding age relatively constant. In the group of recruits from 17 through 22 years of age, complete data were avaiable for 85 men who had finished no more than the 8th grade and for 97 men who had at least graduated from high school.

In the higher education group, 52 per cent of the men liked 16 or more of the activities. In the lower education group, only 32 per cent liked 16 or more items. In the higher educatión group, 86 per cent liked 11 or more items, while in the lower education group, only 74 per cent liked as many items. Table 2 presents the distributions for the two educational groups. Comparison of these differences with those found between the two different age groups reveals that much more variation in range of interests is related to changes in age than to differences in education, although the influence of education is nevertheless marked.

The total group for whom complete data were available was divided into those recruits who had completed the 8th grade or less and those who had completed more grades than this. Table 3 presents the per cent of 17-22 year old men and 29–49 year old men in each of these two educational groups who liked each of the listed items. The figures are also presented for all of the men who had completed no more than the 8th grade and all who had completed more than this.

Comparison of the two columns of figures for the two educational groups shows fewer changes in interests related to education than to age. The more educated men show more interest in dancing, in going on dates, in going to parties and in athletics associated with secondary schools. The less well educated groups like to play horseshoes more than the others. The differences in interests accompanying change in educational status well conform to the opportunities offered by our educational institutions.

In order to determine the relative influence of age and education upon responses to the items, the 22 items were ranked in order of popularity among the 17–22 year old men and among the 29–49 year old men. The correlation between these two rank orders was +.52, indicating substantial shifting of item popularity as one moved from the younger to the older group. The items were then ranked in order of popularity among the men completing 8 grades or less and among the men completing more than 8 grades. The correlation between these two rank orders was .78. These two correlations make plain that, although educational status affects item response, age differences are far more effective.

The author's impression is that age differentials must be allowed for in evaluating range of interests. Educational differences, however, need not be considered if convenience so dictates, although more exact information will be obtained if these differences are also taken into account.

The primary question here, as it was with the oral list, is how well does this list differentiate between people who are making a good adjustment to military training and people who are not.

Data were available for the three groups of non-adjusting individuals mentioned previously. The first group consisted of the 25 recruits selected by psychiatrists as requiring further study not involving intellectual level or literacy. The mean age of this group was 23.1 years with a standard deviation of 6.24 years. The second group consisted of the 23 recruits selected by psychiatrists as requiring examination concerning possible mental deficiency or illiteracy. The mean age of this group was 26.6 years with a standard deviation of 5.76 years. The third group consisted of those 114 cases who were given inaptitude discharges. The mean age of this group was 25.7 years and the mean school grade completed was 8.8.

Distributions of the number of items liked by these different groups are presented in Table 4. Two different groups of normal recruits are presented for comparative purposes.

Table 4

Distributions of Number of Likes Obtained with Printed List for Various Groups of Normal and Non-Adjusting Recruits

No. of Items Liked	Screened for other than M.A. or Reading		Screened for M.A. or Reading		Aptitude Discharges		Normal Recruits not Screened		Normals Matched with Aptitude Discharges	
	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %
22			2	100	1	100	46	100	6	100
21					2	98	34	94	3	94
20	1	100			2	97	45	89	4	92
19	1	96	1	92	1	95	39	83	6	88
18	2	92	1	88	2	94	64	78	11	83
17	2	84	1	82	1	93	38	69	1	74
16	1	76			2	92	45	64	9	73
15			2	78	2	90	42	58	4	65
14			3	70	4	88	54	52	7	61
13	2	72	1	56	3	85	38	45	4	55
12	1	64	1	52	2	82	51	39	8	52
11	4	60	1	48	2	81	47	32	11	45
10	1	44	1	44	4	80	39	26	5	35
9	2	40	4	39	7	75	29	21	11	29
8	2	32	2	22	11	70	28	17	3	21
7	2	24			11	60	32	13	7	18
6			1	13	10	50	16	8	5	12
5	1	16	1	8	11	41	17	6	3	8
4	2	12			14	31	17	4	3	5
3	1	4	1	4	12	10	7	2	3	3
2					10	8				
1							4	1		
Total	25		23		114		732		114	

The first group of 732 normal recruits includes all recruits for whom ages were available and who were not selected by psychiatrists as needing further examination. The mean age of this group was 25.4 years with a standard deviation of 4.0 years. The second group of normal controls consisted of 114 recruits selected from the larger control group and matched with the 114 discharged cases on the basis of age, highest grade completed, and age at leaving school. The effectiveness of this matching is revealed by comparing the mean age of 25.9 of this normal group and their mean grade attainment of 8.8 with the similar figures just presented for the 114 discharged recruits.

Comparisons of the distributions indicate that, with education and age constant, the non-adjusting individuals and the normals are differentiated by their ranges of interests as measured by this list. When a group of individuals is tested and then from this group, people are selected by psychiatrists as requiring further examination, this selected group will have restricted ranges of interests. Similarly, when a group, already studied by psychiatrists and psychologists and found to have an unfavorable prognosis, is tested, people in this group will also have restricted ranges of interests.

The mean scores of the two matched groups were compared. The mean score of the 114 normal cases was 12.8 with a standard deviation of 5.2. The mean score of the 114 discharged cases was 6.9 with a standard deviation of 5.0. The critical ratio of the difference between these two means, based on the standard error of the difference and taking into account the low but positive correlation between the groups, was 9.7. Beyond all doubt the scale differentiates between the two groups.

In the discharged group, 50 per cent liked six or fewer items, while in the normal group, only 12 per cent liked 6 or fewer items. In the normal group, 35 per cent liked more than 15 items while in the discharged group only 10 per cent like as many items. The relatively small amount of overlapping indicates the list measures a trait which clearly differentiates those people who have adjusted to military training and those who have not and this trait can be measured with sufficient exactitude to be useful in individual cases.

Conclusion

The preliminary work with the orally presented list and the later work with the printed list show that the range of interests offers an indirect and convenient method for predicting adjustment to military training. The results show more overlapping between groups with the printed list than with the oral list. The more careful control of age and educational factors obtained in the analysis of the printed list, however, must be considered. This and the greater convenience and objectivity of the printed list give it added advantage.

The method can be useful in several situations. First it may be used as a screening device. As it does not select mental defectives and perhaps many other groups aimed at in psychiatric screening, the list is in itself not a sufficient screening technique. Hunt, Wittson and Harris[4] have concluded that screen tests are most useful to the military psychiatrist in supplementing the results of his clinical interview. Most acceptable is their suggestion that screening tests, such as developed here, and the psychiatric interview are best used to complement each other.

The method is also useful in the collection of information to be used in making the final judgment as to whether a recruit should be retained in military service or not. The validity of these judgments is dependent upon the reliability, the relevancy and the amount of information upon which they are based.

The analysis leaves little question that groups not adjusting well to military training tend to have significantly restricted ranges of interests and this restriction can be measured reliably in individual cases and used in making prognostic judgments.

Received July 17, 1944.

[4] Hunt, W. A., Wittson, C. L., and Harris, H. I. The screen test in military selection. *Psychol. Rev.*, 1944, 51, 37–46.

A Note on the Problem of Brain Damage in Rehabilitation and Personnel Work

Howard F. Hunt

University of Minnesota

Brain injury in the adult human being is usually followed by alterations in personality and intellectual ability as well as by alterations in physical status. The symptomatology and vocational implication of these alterations deserves discussion since many returning brain-injured veterans will be only partially incapacitated and will desire some means of self-support. The question of the employability of brain-injured persons is an important one at the present time, moreover, since this condition is not too uncommon among the general population as a result of accidents, disease, and excessive use of alcohol. This discussion is intended to provide personnel and rehabilitation workers who have not specialized in clinical psychology with useful information for dealing with brain damage cases. Only the psychologic symptoms of greatest practical import from the standpoint of vocational guidance and personnel work will be discussed here. No attempt will be made to discuss the "basic, underlying defect, or defects" of which these more or less superficial symptoms are a function.

The residuals of brain injury may, for purposes of convenience, be classified under two headings: primarily physical or somatic symptoms and symptoms which are primarily behavioral or psychologic. Among the more important physical symptoms are included: weakness and fatigability, convulsive disorders, sensory and motor disorders (including the aphasias), and defects in muscular coordination. These symptoms are generally detectable by routine physical and neurological examinations and from the medical history obtained from such a person. With the possible exception of the aphasias, the somatic disabilities usually come within the purview of the physician and will not be discussed here.

The psychological symptoms consist superficially of intellectual deterioration and alterations in the individual's emotional life and personality. The severity of the symptoms depends largely upon the severity and extent of the damage incurred and upon the prior personality organization and intellectual level of the affected individual, varying quantitatively from exceedingly subtle alterations in persons who superficially seem quite normal to gross dementia with disorientation as to time, place and even person. In some cases both the psychologic and somatic symptoms may be present in varied degrees and combinations, while in other cases either the psychologic or the somatic symptoms alone may be appreciable. Even relatively small areas of damage in certain locations may produce distinct physical symptoms but little or almost no psychologic alteration, whereas lesions placed in other areas of the brain may result in behavioral defects unaccompanied by appreciable somatic abnormality. We are primarily concerned with the latter type of case in this discussion.

The intellectual deterioration or loss of intellectual ability resulting from brain damage apparently involves those more dynamic aspects of intelligence which are crucial for problem solving. Brain-damaged persons commonly show some loss of memory for recent events, a symptom which is related to their decreased ability to learn unfamiliar material or new tasks. They also show rigidity and perseveration in their mode of attack on new problems and in their behavior in general. This defect is manifested in their decreased ability to organize or synthesize new behavior patterns and in their tendency to continue to apply old, familiar methods of solution to new problems even after the method has proved unsuccessful. This characteristic may appear in social situations in the guise of decreased ability to follow rapid shifts in the form and content of conversations and diminished skill in coping with sudden turns of events. In addition to these symptoms, a general slowing of thinking and speed of reaction as well as an increased distractibility are often observed. In general the old, familiar response patterns, habits, and skills are relatively unaffected by mild to moderate degrees of brain damage.

Alterations in the individual's emotionality and, consequently, in personality frequently follow brain damage. Emotional reactions tend to be more easily aroused and to greater heights in the brain-damaged person than in the normal person, and they tend to abate more rapidly once aroused. Thus, these people often give the impression of a relative lack of inhibition or emotional control as well as of deficient foresight and concern about their own future or the consequences of their acts. Many show a mild degree of perhaps unjustified optimism and buoyancy or euphoria, while some others may show a tendency toward an exaggeration of their previous personality characteristics. These symptoms, plus the accompanying intellectual changes, set the stage for childish or antisocial and unethical behavior in predisposed persons. In general, the individual's power of adjustment to society and to the problems of everyday life is diminished.

No specific personality "type" seems to be invariably peculiar to or characteristic of brain-damaged persons. In the case of adults, the defects described above are imposed on a matured personality "system" or structure. The resultant behavior characteristics will thus be a product of the interaction between the individual's original personality and intellectual level and the specific residuals associated with brain damage, plus the frequently exaggerated reactions of this modified personality to the stresses and strains incidental to the demands of daily life. This latter factor may be particularly important in some cases since the brain-injured person's goals and expectations may remain relatively unaltered though his capacity for achieving them is reduced. Thus, a person whose powers of adaptation and emotional control are diminished must often live under and adjust to particularly trying circumstances. It is not surprising that some brain-injured persons develop psychoneuroses or exhibit socially unacceptable behavior. Sensori-motor and speech defects or convulsive disorders may also complicate this situation, as may psychoneurotic reactions on the part of the affected person to the idea of being handicapped or to the circumstances associated with a spectacular, injury-producing accident or illness.

Because there is no regeneration of destroyed neuronal tissue within the brain, the permanence of the defects associated with brain damage has sometimes been assumed. Partial and sometimes almost complete clinical recovery does occur in some cases, however; but in which cases it will occur cannot yet be predicted accurately. Recovery, if it is to occur to an appreciable degree, is usually well under way within a period of a few months after injury and is not generally to be expected in cases of long standing deterioration.

This discussion has been primarily concerned with the symptoms of cases in which brain damage occurs after the individual has attained almost full intellectual growth, that is after the person is fifteen or sixteen years old. The defects resulting from damage earlier in life are somewhat similar with additional complications arising from arrested or uneven intellectual and personality development; consequently defective education, often with marked and specific deficiencies in one or more of the tool subjects such as arithmetic or reading; and the development of various compensatory habits and unacceptable reaction patterns. Probably only the less severely affected cases of this type will be encountered in the employment situation since, in most cases, the segregating and selecting influences of school and society have eliminated the grossly unfit from the active labor market.

The vocational prognosis attached to brain-damaged persons depends upon a number of factors. The severity of the handicap in such persons is relative to their original intellectual level and personality. It depends not only upon the extent and placement of the damage and upon the absolute magnitude of the intellectual and emotional alterations but also upon the individual's original intellectual ability, education and training, and the excellence of his personality integration and adjustment. The magnitude of the handicap is also relative to the type of work he wishes to perform and to his interest in and familiarity with it.

The exact vocational implication of mild to moderate psychologic defect following brain damage is as yet unknown, but a few generalizations may be drawn from our relatively limited clinical experience with it and on the basis of the symptoms of the condition. Occupations placing a premium on physical stamina, inventiveness and ingenuity, emotional control, and diplomacy in inter-personal relations are probably contraindicated in the majority of cases just as are positions involving great personal responsibility both for the welfare of other persons and for vital programs or equipment. Special experience and training or ability and good personal adjustment coupled with relatively minimal intellectual and emotional changes, however, would warrant exception to the above generalizations in some cases. Usually, nevertheless, even mild psychologic defect is probably an unfavorable prognostic sign with regard to both vocational and personal adjustment.

Though diagnosis and treatment of brain damage are primarily medical problems, the standard techniques utilized by physicians fail to yield certain types of information crucial to the vocational management of affected persons. Medical facilities for detecting and estimating the severity of the psychologic symptoms associated with brain damage consist largely of unrefined, uncalibrated tests aided by subjective, clinical impressions and estimates. These techniques yield relatively unreliable estimates of degree of deterioration. Also, since they are relatively insensitive, they may fail to detect minimal defects, especially in cases where the physical symptoms are not appreciable. In a few cases, moreover, the physical and neurologic examination findings may be entirely negative, the psychologic defects minimal, and the brain damage detectable only by refined psychometric methods. Recent advances in clinical psychology have provided more adequate and more sensitive methods. Though these methods do not "measure" in the strictest mathematical sense, they represent a marked improvement over the traditional medical techniques in that they are somewhat more sensitive and do yield rough estimates of degree of deterioration. They are particularly useful in evaluating the employable brain-damage cases—those with relatively minimal defect.

Clinical psychology has, however, no special technique to offer for the detection of minimal alterations in the emotionality of brain-damaged persons. The standard personality tests are of considerable value in assessing the actual personality characteristics of these people, nevertheless. A personality deviation in a brain-damaged person may have more serious implications than a similar deviation in a person who is intact.

To be of practical value in the personnel or vocational guidance situation, a psychometric instrument must not require extended administration time or extensive, cumbersome equipment, and yet it must be maximally sensitive. Ease of administration and scoring as well as interpretability are also crucial. Of course, the test must have clinical validity and be applicable to the type of person being tested, applicable both from the standpoint of its content and from the standpoint of its standardization. As is well known, however, the use of tests provides no royal road to diagnosis. The pitfalls for the unwary are many, and the thoughtless clinician all too easily may be lulled into a false sense of security by the arithmetic form and so-called "objectivity" of test results. A test is no better than the clinician who uses it, and an otherwise useful test may be rendered ineffective by thoughtless, careless administration and naive interpretation. The examiner must be thoroughly familiar with the administration technique as well as fully cognizant of the psychologic aspects of the task and total situation with which the tested person is confronted. This is particularly important in brain-damaged cases because of their characteristic emotionality and psychologic defects. Naturally, test results must be considered as but a part of the evidence required for diagnosis and must always be interpreted in the light of all the rest of the data relevant to the case being studied.

The standard intelligence tests have often been used to detect intellectual deterioration, but they are somewhat ineffective tools for this purpose from the strict psychometric standpoint because of their relative insensitivity to this condition. They may, however, provide the occasion for behavior of great diagnostic significance to an experienced clinical psychologist. In general, the same evaluation applies to the Rorschach (ink blot) method when used as a deterioration test. Moreover, much specialized experience and training are required for proper administration, scoring, and interpretation of this procedure. Accordingly, the use of psychometric instruments specially designed for the detection of deterioration is advised, particularly in borderline cases.

A number of deterioration tests have been developed, but the more effective of such tests have been based upon what might be called the differential score technique which was originally suggested by Babcock (1). This technique involves measuring the discrepancy between a person's performance on tests sensitive to brain damage (tests of learning, memory, abstract thinking) and his performance on tests relatively insensitive to this condition (tests of vocabulary, information). Significant discrepancies indicating impaired performance on the sensitive tests relative to performance on the insensitive tests would thus suggest brain damage. This neat situation is complicated, however, by the tendency of depressed, anxious, schizophrenic, and other varieties of disturbed persons to obtain similarly discrepant scores on the two types of tests. Thus, such pathologic discrepancies, or deterioration scores, are ambiguous diagnostically and should be confirmed by case history and neurologic examination data before anything stronger than a suspicion of brain damage is justified.

In cases with pathologic deterioration test scores, differentiation as to the etiology of the score is important from the standpoint of long-time vocational prognosis. Pathologic scores attributable to emotional-motivational disturbances as in the depressions and the anxiety states indicate an impairment of intellectual efficiency which usually disappears when the afflicted individuals recover. Schizophrenic deterioration generally disappears in a like manner in those cases which do recover. In contrast, the intellectual impairment associated with brain damage is usually considerably less transient and is accompanied, therefore, by a less favorable vocational prognosis. Because of the proneness of brain-damaged persons to emotional disturbances, the more permanent impairment associated with the brain damage may be temporarily augmented by depression, anxiety, and the like. Under these conditions, the ultimate intellectual status of the person is difficult to determine since some of the impairment indicated by the test score may be expected, in many cases, *to disappear when emotional equilibrium is regained.*

Deterioration test scores are thus not a final index of vocational prognosis but rather a diagnostic and prognostic aid. The extent to which they aid diagnosis and prognosis depends to a substantial degree upon the skill and clinical acuity of the interpreting clinician.

Of the available, specially devised deterioration tests for adults, the Shipley-Hartford Deterioration Test (8) is the easiest to administer. It is a twenty minute, self-administering, paper and pencil test which may be used only with cooperative, test-sophisticated subjects of average or above average intelligence.

The Babcock-Levy Revised Test for Efficiency of Mental Functioning (2) is an individual test very similar to the standard individual intelligence tests in the time, skill, and amount of equipment required for administration. It is applicable to cooperative persons of a somewhat less restricted range of intelligence and test-sophistication than is the Shipley. The norms for these two tests include no correction for normal deterioration with advanced age, so subjective correction for this factor must be made in the interpretation of the test performances of older persons.

The Hunt-Minnesota Test for Organic Brain Damage (4) is an individual test requiring some practice but relatively little equipment for administration. The test can be completed in from fifteen to thirty five minutes, depending upon whether the short or the long form is used. It is applicable to cooperative adults with mental ages of over eight years, and the norms include a statistical correction for normal age changes within the range of sixteen to seventy years. Scores on this test obtained from persons near the extremes of the age and intelligence ranges must be interpreted with considerable caution, however.

"Psychological deficit" has been suggested (6) as an inclusive but neutral term for the impairment of intellectual efficiency associated both with brain damage and with emotional-motivational disorders. The three tests described above are apparently fairly effective in detecting this deficit. In the development of the Hunt test, an attempt was made to provide a special means for identifying those pathologic scores attributable to emotional-motivational disturbances so that the test would then be a specific test for the deterioration associated with brain damage. This attempt has been only partially successful. Since neither the Shipley nor the Babcock tests make this discrimination, pathologic scores on any of these three tests may indicate either brain damage or an emotional-motivational disturbance accompanied by poor cooperation, motivation, and attention. Consequently, great care must be exercised in the interpretation of scores obtained on these tests.

Received July 18, 1944.

References

1. Babcock, H. An experiment in the measurement of intellectual deterioration. *Arch. Psychol.*, N. Y., 1930, No. 117.
2. Babcock, H., and Levy, L. *Test and manual of directions: the revised examination for the measurement of the efficiency of mental functioning.* Chicago: C. H. Stoelting Co., 1940.
3. Hunt, Howard F. A practical, clinical test for organic brain damage. *J. appl. Psychol.*, 1943, 27, 375–386.
4. Hunt, Howard F. *The Hunt-Minnesota test for organic brain damage.* Minneapolis: The University of Minnesota Press, 1943.
5. Hunt, Howard F. A note on the clinical use of the Hunt-Minnesota test for organic brain damage. *J. appl. Psychol.*, 1944, 28, 175–178.
6. Hunt, J. McV. *Personality and the behavior disorders.* New York: The Ronald Press, 1944.
7. Pollack, B. The validity of the Shipley-Hartford Retreat test for "deterioration." *Psychiat. Quart.*, 1942, 16, 119–131.
8. Shipley, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, 9, 371–377.
9. Shipley, W. C., and Burlingame, C. C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *Amer. J. Psychiat.*, 1941, 97, 1313–1324.

Personality Patterns of Adolescent Girls:
II. Delinquents and Non-Delinquents

Dora F. Capwell

Trainee Acceptance Center, Public Schools, Pittsburgh, Pa.

In a study devoted to the personality patterns of adolescent girls who show improvement in IQ (2) two groups of girls were used as subjects, a group of delinquents and a group of non-delinquents. A series of personality tests were administered to each group on two occasions along with tests of intelligence and academic achievement. The results of the personality tests, aside from their contribution to the original problem, are of special interest due to the degree and manner in which they differentiated the delinquents from the non-delinquents. Most present day students of delinquency agree with the point of view expressed by Lowrey (10). "Delinquency is probably most frequently due to the subtle effects of interactions between individual and environment, leading to the establishment of particular personality sets." However, there has been disagreement regarding the success with which one can measure and delineate these personality sets by means of standard, objective tests of personality. The following results show the differences of personality between a group of delinquent girls and a group of non-delinquent girls as measured by one series of personality tests.

The procedure has been described in detail in a previous report (2). A total of 101 delinquent girls at the Minnesota State School for Girls and 85 non-delinquents in the Public Schools of Sauk Centre, Minn., were given a psychological examination and re-examined from 4 to 15 months later. The personality tests which were given twice were the Minnesota Multiphasic Personality Inventory (6), the Washburne Social Adjustment Inventory (13) and the Pressey Interest-Attitude Test (11). Two other tests of personality, the Terman-Miles Test of Masculinity-Feminity (12) and the Vineland Social Maturity Scale (3), were given just once. The levels of intelligence and academic achievement were determined by the Kuhlmann Tests of Mental Development (9) and the Stanford Achievement Test (8).

Results

Before examining the personality test results, it is important to note the differences between the two groups in intelligence and achievement.

The non-delinquents were girls of higher intelligence, showing a mean IQ on the first test of 101, as compared with a mean of 87 for the delinquents. The standard deviation for each group was 17. Each group showed some retardation in school achievement when compared with the norms of the Stanford Achievement Test. The delinquents were more retarded than the non-delinquents, as would be expected, but the differences are not statistically reliable. The amount of retardation in terms of grade scores is shown in Table 1.

Table 1

Difference between Actual Grade Level and Achievement Grade Score *

	Delinquents $N = 97$**		Non-Delinquents $N = 85$		$\bar{D}/\sigma D$ between Delinquents and Non-Delinquents
Score	Mean Diff.	S.D.	Mean Diff.	S.D.	
Total Score	−1.30	1.41	−.82	1.50	2.28
Reading Score	−.55	1.52	−.49	1.46	.28
Arithmetic Score	−1.96	1.76	−1.28	2.03	2.42

* 1.00 equals one total grade.

** Four delinquents were absent when these tests were given.

The personality tests discriminated the delinquents from the non-delinquents with varying success; two of them showed striking differences. The *significance of differences between the two groups on the Minnesota* Multiphasic Personality Inventory is shown in Table 2.

Table 2

Minnesota Multiphasic Personality Inventory—Significance of Difference of Raw Scores

	Difference between Delinquents and Non-Delinquents	
Scale	First Test $D/\sigma D$	Second Test $D/\sigma D$
"?"	2.93	4.25
L	1.75	3.30
F	7.21	5.95
Hs	3.11	3.12
D	4.59	2.92
Hy	2.74	.00
Pd	16.00	14.00
Pa	12.00	8.03
Pt	6.64	7.36
Sc	7.10	8.55
Ma	8.00	7.95

Each scale except the Hy, or Hysteria scale, shows a clear differentiation between the two groups. The greatest difference appears in the scores for Pd (Psychopathic Deviate). The extent to which the Multiphasic differentiated the groups may be seen in Table 3, and Figures 1 and 2 show graphically the differences in average T-scores between the two groups.

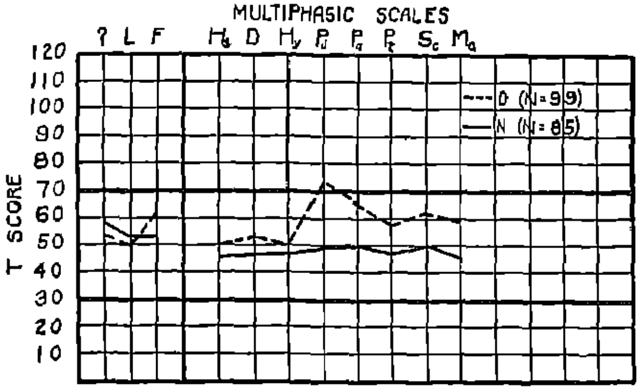


FIG. 1. T-score profile on the Multiphasic Inventory for the Delinquent (D) and Non-Delinquent (N) groups, 1st test.

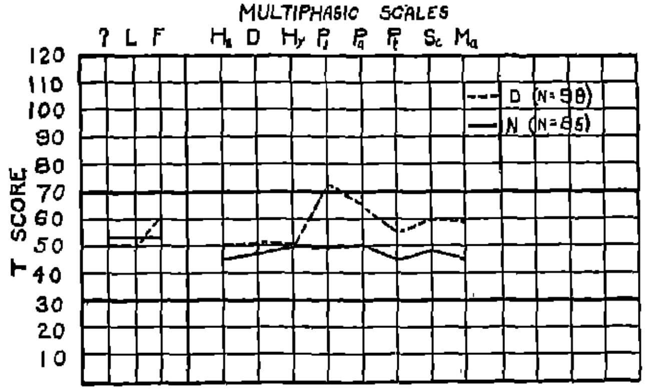


FIG. 2. T-score profile on the Multiphasic Inventory for the Delinquent (D) and Non-Delinquent (N) groups, 2nd test.

Table 4 shows the significance of differences between the raw scores of the other personality tests. The Washburne and the Vineland differentiated between the delinquents and non-delinquents, but the Pressey Interest-Attitude and the Terman-Miles did not. On the Washburne 54% of the delinquents reached or exceeded the 75th percentile of the non-delinquents' scores on the first examination, and 51% did so on the

Table 3

Percentage of Overlap on the Multiphasic

Multiphasic Scale	Percentage of Delinquents who Reached or Exceeded the 75th Percentile of Non-Delinquents: 1st Test	Percentage of Delinquents who Reached or Exceeded the 75th Percentile of Non-Delinquents: 2nd Test
Hs	41%	48%
D	48%	37%
Hy	32%	19%
Pd	93%	92%
Pa	84%	65%
Pt	61%	61%
Sc	64%	71%
Ma	64%	71%

second examination. The Vineland showed differences in the other direction, in that on the other personality tests a high score tends toward maladjustment, while on the social maturity scale a high score is a favorable one. On the social maturity scale only 31% of the delinquents reached or exceeded the median of the non-delinquents.

Table 4

Other Personality Tests—Significance of Difference of Raw Scores

Test	Test No.	Difference between Delinquents and Non-Delinquents D/σD
Washburne	1st	5.31
Washburne	2nd	4.54
Pressey	1st	.62
Pressey	2nd	.61
Terman-Miles	(one only)	.80
Vineland *	(one only)	2.96

* D/σD between social quotients.

Inasmuch as the two groups which were differentiated by the Multiphasic, the Washburne, and the Vineland differ in level of intelligence as well as in regard to delinquency, it is necessary to find out if the personality test differences are related to the differences in intelligence. Hence, 52 delinquents from the group were matched within two IQ points with 52 of the non-delinquents in order to compare personality differences and percentage of overlap on those tests which differentiated the groups reliably. The mean IQ of each group was 95 with a standard deviation of 14. Only those tests were compared which differentiated the larger groups. Results obtained with the matched groups may be seen in Table 5.

The Multiphasic continued to differentiate the delinquents and non-delinquents reliably with the exception of the Hs scale. Hy was not used because it did not differentiate the larger groups. The Washburne fell slightly below satisfactory reliability of differences, and the Vineland failed to demonstrate any real difference when the groups were matched for IQ. Hence, the differences on the Vineland appear to be related more to intelligence than to delinquency, but the differences on the Washburne and the Multiphasic continue to be significant.

Table 5

Matched Groups of Delinquents and Non-Delinquents: Significance of Difference of Raw Scores on First Examination and Percentage of Overlap

Test	$D/\sigma D$	Percentage of Delinquents who Reached or Exceeded 75th Percentile of Non-Delinquents
Multiphasic		
Hs	1.95	—
D	3.91	55%
Pd	10.34	93%
Pa	8.43	84%
Pt	4.86	57%
Sc	4.06	57%
Ma	5.22	63%
Washburne	2.70	50%
Vineland	.38	—

Discussion

The test which differentiated most clearly between these two groups was the Multiphasic. The results, which include the three validating scores, show first of all that even quite young adolescents do answer a well-devised personality inventory validly. The mean validity scores are close to a T-score of 50 for all groups, except for the delinquent scores on the F-scale; these were higher, but not sufficiently so as to invalidate the results. They would be expected to be somewhat higher than the F-scores for the non-delinquents because of the much greater amount of maladjustment in the delinquent group, and since, as stated before, there is a slight tendency for the F-score to be higher as gross maladjustment increases. The non-delinquents had many more "?" responses than the delinquents, due probably to two factors: first, they were a slightly younger group and found more items they did not understand, and secondly, the delinquents had more leisure at the time of taking the tests. On the whole, however, the results from both groups showed a reassuring degree of validity.

It has long been believed that delinquents are more generally unstable than the normal population, and the Multiphasic Inventory bore out this belief by showing a significant difference of mean scores between delinquents and non-delinquents on seven of the eight scales, the delinquents scoring further away from the mean for the normal population. Even when the scores were not equivalent to a T-score of 70 or over, the criterion for significant maladjustment, they still were further toward the maladjustment end of the scale than were the scores of the non-delinquents. The most conspicuous differences were on the scales of Psychopathic Deviate and Paranoia. The delinquents had a mean T-score of 73 on the Pd scale and 65 on the Pa scale.

The delinquents were not as depressed as one would imagine, on their arrival at a correctional institution. That they are not is a finding which is consistent with the theory of the psychopathic deviate set forth by Hathaway (6), suggesting that the psychopathic inferior or deviate has trouble in his relations to society partly because he does not react with the emotions with which the normal person reacts under similar circumstances. Therefore, although the newly committed delinquent may seem to feel very badly, cry noisily, and appear to be extremely depressed, the test scores and also the speed with which she often recovers from this mood, attest to the belief that it is a relatively superficial emotion. In the non-delinquent group only two girls made scores on the Pd scale as high as the mean for the delinquent group. Of these two one was a girl so maladjusted that she was being considered for foster-home placement by the child welfare worker of the county, and the other was a girl who had left school for a time, returned of her own volition, but was again considering leaving and was so upset that she was of concern to the school principal; she herself asked for a conference with the examiner to talk about her difficulties. Thus, this test measures adjustment in the same sense in which non-delinquents are better adjusted than delinquents.

Two other personality tests, the Pressey Interest-Attitude Test and the Terman-Miles Test of Masculinity-Femininity, when their raw scores were compared as with the other tests, did not discriminate between the delinquents and non-delinquents. This is contrary to the results implied by the report of Durea and Fertman (5), who gave the Pressey to 180 delinquent girls. Their study merely showed, however, that the scores from the delinquents compare "unfavorably with norms for non-delinquents." No control group was used at the same time. In the present data the delinquents compare unfavorably with the *norms*, but so does this particular sample of normal cases. These same normal girls made normally expected scores on the other tests. The Pressey apparently does not measure the type of adjustment measured by the Multiphasic or the Washburne. The Terman-Miles did not show significant differences between the groups, but each group contained some cases of extreme scores, as one may judge from the standard deviation of the groups. The deviate cases were not always the girls who tested as most maladjusted on other tests.

The Vineland Social Maturity Scale did not show as much difference in social maturity as might be expected between the delinquents and non-delinquents. These girls do not show as much social retardation as the delinquent boys reported by Doll and Fitch (4), but there are other marked differences between the boys they studied and the girls of this study, the most important one being that the boys were much more retarded mentally than the girls are. The median mental age for the boys was 9.3 years, as against a median life age of 14 years. The social quotients were nearer the mental ages than the life ages. In the present group of delinquent girls the mean social quotient is 4 points below the mean IQ and is more related to intelligence level than delinquency.

In concluding the discussion of these test results a word may be said about the use of personality tests as measures of adjustment both with groups and individuals. Whether or not one considers them valid and helpful instruments to aid diagnosis appears to depend largely on the selection of tests one uses. Boynton and Walsworth (1) used six tests, all different from the ones used in the present study, with 47 delinquent and 50 normal girls and found that only one score of one test yielded a C.R. of 3.00 between the groups. They concluded that the tests "in the main do not provide empirical evidence of sufficient validity to justify one in putting a great deal of faith in them in individual and group diagnosis." They concluded further that since there was such disparity between the test results, delinquent behavior is not necessarily associated with personality aberrations. Somewhat different conclusions are warranted from the results of tests used in the present study. Two of them, the Terman-Miles M.-F. Test and the Pressey Interest-Attitude Test, were not helpful in differentiating delinquents and non-delinquents. Two others, however, gave satisfactory differences and should be helpful in either individual or group diagnosis; these two were the Minnesota Multiphasic Personality Inventory and the Washburne Social Adjustment Inventory. The results of both suggest that personality aberrations frequently *are* associated with delinquency. The use of personality tests with any special group necessitates careful selection of tests on the basis of the way they were standardized as well as their suitability for providing the desired information. Beyond that, experience with the results in the particular group concerned is necessary before a fair evaluation can be made of the usefulness of the test. Generalization from a selected group of personality tests to all tests of that type is not warranted any more than from one intelligence test to another.

Summary and Conclusions

A group of 101 delinquent and 85 non-delinquent girls were tested with a battery of personality tests and retested from 4 to 15 months later. Differentiation between the two groups was measured by computing the significance of difference of the mean scores for each group and also the percentage of overlap. The groups differed in level of intelligence as well as in delinquent tendencies, so the effect of mental level on the personality test scores was investigated by similar statistical treatment of the scores of 52 girls from each group who were matched for IQ. The results led to the following conclusions:

1. The Minnesota Multiphasic Personality Inventory and the Washburne S.-A. Inventory discriminated the delinquents from the non-delinquents in degree of personality adjustment.
2. The Vineland Social Maturity Scale showed differences which were more related to intelligence than to delinquency.
3. The Pressey Interest-Attitude Test and the Terman-Miles Test of Masculinity-Femininity did not discriminate the delinquents and non-delinquents.
4. Personality tests may be of value with individuals or groups in measuring and describing the personality patterns of delinquent as distinguished from non-delinquent girls.

Received June 12, 1944.

References

1. Boynton, P. L., and Walsworth, B. M. Emotionality test scores of delinquent and non-delinquent girls. *J. abnorm. soc. Psychol.*, 1943, 38, 87–92.
2. Capwell, Dora F. Personality patterns of adolescent girls: I. Girls who show improvement in I.Q. *J. appl. Psychol.*, 1945, 29, 212–228.
3. Doll, E. A. A genetic scale of social maturity. *Amer. J. Orthopsychiat.*, 1935, 5, 180–188.
4. Doll, E. A., and Fitch, K. A. Social competence of juvenile delinquents. *J. of crim. Law Criminol.*, 1939, 30, 52–67.
5. Durea, M. A., and Fertman, M. H. Emotional maturity of delinquent girls. *Amer. J. Orthopsychiat.*, 1941, 11, 335–338.
6. Hathaway, S. R. The personality inventory as an aid in the diagnosis of psychopathic inferiors. *J. consult. Psychol.*, 1939, 3, 112–117.
7. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *J. Psychol.*, 1940, 10, 249–254.

8. Kelley, T. L., Ruch, G. M., and Terman, L. M. *Stanford achievement Tests: Manual of directions.* Yonkers-on-Hudson, New York: World Book Company, 1940.
9. Kuhlmann, F. *Tests of mental development.* Minneapolis, Minn.: Educ. Test Bureau, 1939.
10. Lowrey, L. G. Delinquent and criminal personalities. In J. McV. Hunt (Ed.). *Personality and the behavior disorders.* New York: The Ronald Press Company, 1944, Vol. II, pp. 794–821.
11. Pressey, S. L., and Pressey, L. C. Development of the interest-attitude tests. *J. appl. Psychol.*, 1933, 17, 1–16.
12. Terman, L. M., and Miles, C. C. *Sex and personality.* New York: McGraw-Hill, 1936.
13. Washburne, J. N. A test of social adjustment. *J. appl. Psychol.*, 1938, 19, No. 2, 125–144.

The Construction of the Guilford-Martin Inventory of Factors G-A-M-I-N

Howard G. Martin

Los Angeles, California

Factor analysis methods have been applied to the problem of isolating independent variables of temperament with the result that several unitary traits have been identified (3, 4, 5) and a series of inventories constructed which attempt to measure some of these traits (1, 6).

The Guilford-Martin Inventory of Factors G–A–M–I–N adds five more temperament variables to the eight already covered by the two preceding tests of the series. The five traits included in the new inventory are: **G**, General pressure for overt activity; **A**, Ascendancy in social situations as opposed to submissiveness; leadership qualities; **M**, Masculinity of attitudes and interests as opposed to femininity; **I**, Lack of inferiority feelings; self-confidence; and **N**, Lack of nervous tenseness and irritability.

Traits G, M, and N were discovered in factor analyses by Guilford and Guilford (3, 5) and trait I by Mosier (7). Traits A, M, I, and N have been recognized by clinical psychologists, though not always defined in the same manner. The definitions of the trait names as used in this connection are operational; that is, in terms of factorial studies and subsequent item analyses.

More than 300 questionnaire items were constructed which were believed to cover the areas of behavior constituting these five traits. This list, stated in question form to be answered by either "Yes," "?," or "No," was administered to 250 men and 250 women college students between the ages of 19 and 30, ranging from sophomores to seniors in southern California colleges and universities.

The items which had been shown by the factor analyses and two previous item analyses to have heavy loadings in a trait were included in the preliminary scoring key for that trait. Typical items from these keys are listed below:

For trait G:

> Are you inclined to be quick in your actions?
> Can you turn out a large amount of work in a short time?
> Are you inclined to rush from one activity to another without pausing for rest?

For trait A:

Do you usually speak out in a meeting to oppose someone who you feel sure is wrong?

Do you find it difficult to get rid of a salesman to whom you do not care to listen or give your time?

Have you ever, on your own initiative, organized a club or group of any kind?

For trait M

Do you like love scenes in a movie or a play?

Do you (or would you) like to go hunting with a rifle for wild game?

Are you disgusted at the sound of foul language?

For trait I:

Do you feel that the average person has made a better adjustment to life than you have?

Do you feel confident that you can cope with most situations that you will meet in the future?

Do you suffer keenly from feelings of inferiority?

For trait N:

Do you often become irritated over little annoyances?

Do you have nervous habits such as chewing your pencil or biting your fingernails?

Do you feel compelled to change your bodily posture frequently while sitting?

Four hundred of the questionnaires (200 males and 200 females) were scored with the preliminary scoring keys thus constructed. The highest 100 and the lowest 100 cases (extreme quarters of the distributions of scores) were used as criterion groups in the item analyses for factors G, A, I, and N. For the item analysis of factor M, the criterion groups consisted of the 100 male cases highest on the preliminary scoring key for the M factor and the 100 female cases lowest on the key.

Scoring weights were found for each response to each item by using Guilford's *abac* method (2). This procedure yielded final scoring keys consisting of 41 items for trait G, 50 items for trait A, 52 items for trait M, 64 items for trait I, and 68 items for trait N. Only nine items were scored for more than one trait.

Table 1

Intercorrelations between Scores of 100 Cases

	G	A	M	I
A	.51			
M	.10	.34		
I	.30	.54	.43	
N	−.27	.16	.34	.70

As a check on the reliability of scoring, the test papers of the remaining 100 cases, including 50 males and 50 females, were scored with these final scoring keys, which, for the purpose, were divided into random halves of items. Pearsonian coefficients of correlation were computed between the scores on these halves of the scoring keys and, when corrected by the Spearman-Brown formula, they became .89 for trait G, .88 for trait A, .85 for trait M, .91 for trait I, and .89 for trait N.

Intercorrelations between the scores derived by means of these scoring keys are shown in Table 1.

The validity of trait M was checked by comparing the distributions on the trait scores of the 50 males and the 50 females which were not used in the item analyses. Ninety-two per cent (46 out of 50) of the males were above the median of the distribution of the 100 scores of the two sexes combined. Ninety-two per cent (46 out of 50) of the females were below the median of this distribution. The overlap was only 8 per cent of the 100 scores. Some males are more feminine in their reactions than some females and vice versa (8). That this difference in some cases may have a glandular condition accompanying it is demonstrated by the fact that one of the males who fell below the median on trait M was found to have a history of hypogonadal function and infantile genitalia.

These types constitute an undetermined percentage of the overlap in scores between the sexes in general. Therefore, the validity coefficient (phi) of .84 for trait M which was computed from these figures is as high as can be expected in view of the fallibility of the criterion.

Received July 28, 1944.

References

1. Guilford, J. P. *An inventory of factors S-T-D-C-R.* Beverly Hills, California: Sheridan Supply Company, 1940.
2. Guilford, J. P. A simple scoring weight for test items and its reliability. *Psychometrika*, 1941, 4, 367–374.
3. Guilford, J. P., and Guilford, R. B. Personality factors S, E, and M, and their measurement. *J. Psychol.*, 1936, 2, 107–127.
4. Guilford, J. P., and Guilford, R. B. Personality factors D, R, T, and A. *J. Abn. & Soc. Psychol.*, 1939, 34, 21–36.
5. Guilford, J. P., and Guilford, R. B. Personality factors N and GD. *J. Abn. & Soc. Psychol.*, 1939, 34, 239–248.
6. Guilford, J. P., and Martin, H. G. *The Guilford-Martin personnel inventory.* Beverly Hills, California: Sheridan Supply Company, 1943.
7. Mosier, C. I. A factor analysis of certain neurotic tendencies. *Psychometrika*, 1937, 2, 283–287.
8. Terman, L. M., and Miles, C. C. *Sex and temperament: studies in masculinity and femininity.* New York: McGraw-Hill Book Company, 1936.

Measuring Progress in Radio Training

Gordon L. Macdonald

New York University

In New York City the War Department in conjunction with the New York State Department of Education set up a school in 1942 to train radio technicians where the instruction was administered by various universities and radio schools. The course of instruction, lasting six months, was divided into two sections of three months each. It was within this educational framework that the problem for this study evolved.

This study grew out of the need for an adequate test to measure the progress of trainees under instruction. Specifically the problem is twofold: (1) to develop an adequate instrument for measuring progress in training that would be valid and reliable at various levels; and (2) to apply the instrument to groups of trainees for the purpose of evaluating their progress and comparing results in the several institutions.

This investigation included 1,156 subjects studying the theory and practice of radio servicing (how to repair radios). These subjects represent three levels of general education: (1) High School Students, consisting of three groups of 28, 34 and 25 subjects between the ages of 16 and 20 studying in a public vocational high school (These students were not in any way involved in the War Department training program and were included in this study for comparative purposes); (2) High School Graduates, consisting of four adult groups of 126, 289, 326, and 216 students between the ages of 18 and 40. This age limit was set by the War Department for radio trainees accepted for a course of study in a commercial radio school under the supervision of the War Department and the New York State Department of Education; and (3) College Students of which there were three groups of 25, 30, and 34, with a mean age of 20 years, studying at three universities in New York State where their courses were limited to three months.

The subjects comprising the two educational levels of adult High School Graduates and College Students, who were trainees of the Signal Corps, War Department, were chosen on the basis of a Civil Service aptitude test and a physical examination by army doctors, and they were further required to join the enlisted reserve of the Army of the United States. The group of College Students were further selected on the basis of their knowledge of higher mathematics and college physics.

There was no standardized objective achievement test in the field of radio available. However, an unstandardized test was discovered consisting of two forms, A and B, each containing 116 items and it was used and standardized in this investigation. It was a paper and pencil, multiple choice, self administering, group achievement test with no time limit, but requiring from one hour and a half to over two hours for completion.

The preliminary standardization was done at the level of High School Graduates of general education, and at two levels of specialized radio training. Group V with three months' training, and Group VII with six months' training were used in this phase of the study.

As a result of these analyses two new forms of the test, revised Form A and Form B, were prepared consisting of the best 99 items from the 116 of each original form. It is these forms which were used in the rest of this investigation.

Training Progress Measured

The revised forms were administered to various groups at three distinct levels of general education to find out if progress in specialized radio training could be measured equally well with the test regardless of background. Norms in radio training thereby can be made available taking into account any contribution of general education.

High School Students (Groups I, II, and III): The subjects at this level of general education had completed the 2½ years of general instruction of a public vocational high school and were in attendance on one of the last three terms of specialized radio instruction. They were tested at the end of 315 hours of radio instruction, or one term (Group I), at the end of 630 hours of radio instruction, or two terms (Group II), and at the end of 945 hours of radio instruction, or graduation (Group III). Both Forms A and B, with some exceptions, were administered to all members of these groups.

The means, sigmas of distribution, and C.R.'s between the means on both forms, for the three groups, are given in Table 1.

The levels of specialized radio training are differentiated one from the other by both forms of the test as is indicated by the difference between the means; and these differences are significant as indicated by the critical ratios.

High School Graduates (Groups IV, V, VI, and VII): The adult radio trainees who comprised this level of general education were all high school graduates. The groups were divided according to specialized training in radio as follows: Group IV being novices with no training; Groups V and VI being students with three months' full-time training of 624 hours; and Group VII having completed the six months' full-time course of training of 1,248 hours.

Table 1

Progress in Specialized Radio Training at the General Educational Level of High School Students, Showing M, σ (dist.), for Each Level, and C.R.'s between Levels of Specialized Training Separately for Forms A and B

Form A

Training Period	Group	Mean	σ (dist.)	N
One term (315 hours)	I	43.00	7.77	28
Two terms (630 hours)	II	49.80	11.61	34
Three terms (945 hours)	III	60.92	10.52	25

Critical Ratios

One term (315 hours) vs. two terms (630 hours)	2.75
One term (315 hours) vs. three terms (945 hours)	7.00
Two terms (630 hours) vs. three terms (945 hours)	3.81

Form B

Training Period	Group	Mean	σ (dist.)	N
One term (315 hours)	I	39.00	5.70	28
Two terms (630 hours)	II	46.60	10.17	34
Three terms (945 hours)	III	61.95	9.03	25

Critical Ratios

One term (315 hours) vs. two terms (630 hours)	3.70
One term (315 hours) vs. three terms (945 hours)	10.93
Two terms (630 hours) vs. three terms (945 hours)	6.10

The means, sigmas of distribution, and C.R.'s between the means for all groups on both forms are given in Table 2.

It is evident from the means given in Table 2 that the three levels of specialized radio training of "no training," "three months' training," and "six months' training" are distinguished one from the other. The critical ratios indicate that the differences between the means of the two groups with three months' training (Groups V and VI) are statistically insignificant, while the differences between the means for the other levels of specialized training are highly significant, these values holding for both forms of the test.

The results with Form A (Table 2) show that the difference between the means of the two groups tested at the end of three months of specialized radio training (Groups V and VI) is statistically insignificant, the C.R. being 1.36. The C.R. between the three months' training level (Group V) and the six months' training level (Group VII) is 10.57, and the C.R. between the three months' training level (Group VI) and the six months' training level (Group VII) is 9.39. Thus Form A distinguishes clearly between levels of specialized training in radio among high school graduates.

The results for Form B are similar. The novices, or those with no training (Group IV) are distinguished from the others quite significantly with C.R.'s of 14.08 and 14.74 in comparison with the two groups with three months' training (Groups V and VI); and with a C.R. of 20.12 between them and those at the six months' training level (Group VII). The C.R. of the means of the two groups with three months' training (Groups V and VI) is statistically insignificant, as on Form A, while the C.R.'s of the means between these same groups at the three months' level

Table 2

Progress in Specialized Radio Training at the General Educational Level of High School Graduation, Showing M, σ (dist.), for Each Level, and C.R.'s between Levels of Specialized Training Separately for Forms A and B

Form A

Training Period	Group	Mean	σ (dist.)	N
Three months	V	48.80	9.66	144
Three months	VI	50.35	10.32	162
Six months	VII	62.65	13.43	107

Critical Ratios

Three months (624 hours) Group V	vs.	Three months (624 hours) Group VI	1.36
Three months (624 hours) Group V	vs.	Six months (1,248 hours) Group VII	10.57
Three months (624 hours) Group VI	vs.	Six months (1,248 hours) Group VII	9.39

Form B

Training Period	Group	Mean	σ (dist.)	N
No training	IV	27.30	12.25	126
Three months	V	46.60	9.92	145
Three months	VI	47.35	10.33	164
Six months	VII	60.30	12.77	109

Critical Ratios

No training Group IV	vs.	Three months (624 hours) Group V	14.08
No training Group IV	vs.	Three months (624 hours) Group VI	14.74
No training Group IV	vs.	Six months (1,248 hours) Group VII	20.12
Three months (624 hours) Group V	vs.	Three months (624 hours) Group VI	.64
Three months (624 hours) Group V	vs.	Six months (1,248 hours) Group VII	9.26
Three months (624 hours) Group VI	vs.	Six months (1,248 hours) Group VII	8.87

Table 3

Progress in Specialized Radio Training at the General Educational Level of College Students, Showing M, σ (dist.), for Each Level, and C.R.'s between Levels of Specialized Training Separately for Forms A and B

Form A

Training Period	Group	Mean	σ (dist.)	N
Two months (416 hours)	IX	65.12	6.03	28
Three months (624 hours)	VIII	74.12	9.39	25
Three months (624 hours)	IX	75.00	9.51	30
Three months (624 hours)	X	76.09	12.00	34

Critical Ratios

Two months (416 hours) Group IX	vs.	Three months (624 hours) Group VIII	3.27
Two months (416 hours) Group IX	vs.	Three months (624 hours) Group IX	4.18
Two months (416 hours) Group IX	vs.	Three months (624 hours) Group X	3.92
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group IX	.54
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group X	.71
Three months (624 hours) Group IX	vs.	Three months (624 hours) Group X	.22

Form B

Training Period	Group	Mean	σ (dist.)	N
One month (208 hours)	IX	31.34	8.13	28
Three months (624 hours)	VIII	69.56	8.73	25
Three months (624 hours)	IX	75.10	10.05	30
Three months (624 hours)	X	73.54	12.96	34

Critical Ratios

One month (208 hours) Group IX	vs.	Three months (624 hours) Group VIII	16.48
One month (208 hours) Group IX	vs.	Three months (624 hours) Group IX	18.38
One month (208 hours) Group IX	vs.	Three months (624 hours) Group X	15.63
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group IX	2.05
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group X	1.40
Three months (624 hours) Group IX	vs.	Three months (624 hours) Group X	.54

of training and that of the group with six months' training (Group VII) are statistically significant, being respectively 9.26 and 8.87. Thus specialized training levels are clearly differentiated by both forms at the general level of education represented by High School Graduates.

College Students (Groups VIII, IX, and X): The radio trainees comprising this level of general education received their specialized radio training at three universities in New York State, the course lasting for three months or 624 hours of instruction. The groups are distinguished from each other according to the university which they attended. Group

IX was tested at the end of one month's training with Form B, at the end of two months' training with Form A, and at the end of three months' training with both Form A and B. Groups VIII and X were tested at the end of three months' training with both forms.

The means and sigmas of distribution of both forms at the different levels of specialized training, and the C.R.'s indicating the statistical significance of the difference between levels of training are given in Table 3.

The means indicate that the levels of specialized training at one, two, and three months are differentiated from each other. The differences on Form A at two and three months' training are statistically significant: a C.R. of 3.27 existing between Group IX and Group VIII; one of 4.18 for Group IX; and one of 3.92 between Group IX and Group X.

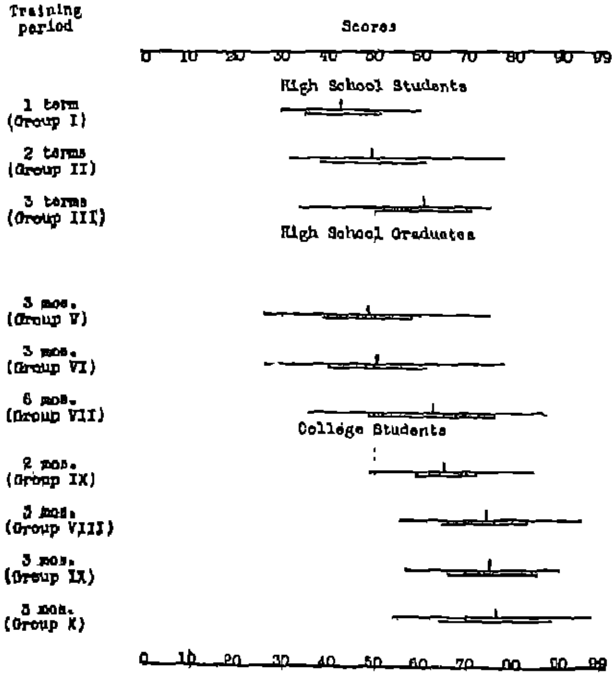


FIG. 1. Levels of radio training, showing M, one σ, and range in scores on Form A for various amounts of radio training classified according to levels of general education.

The data for Form B show that the differences between means at one month's training (Group IX) and three months' training (Groups VIII, IX, and X) are highly significant, the C.R. being 16.48 between Group IX and Group VIII; 18.38 for Group IX; and 15.63 between Group IX and Group X.

The C.R.'s of the means for the groups compared at the three months' level of training are statistically insignificant in all cases. This statistical insignificance of the differences between the means coupled above with the evidence of reliable differences between different levels of specialized radio training, indicates that the test is measuring progress in training.

Progress in radio training is being measured within different levels of general education. Differences between levels of specialized training hold for both forms of the test. A graphic representation of these results is given in Figures 1 and 2.

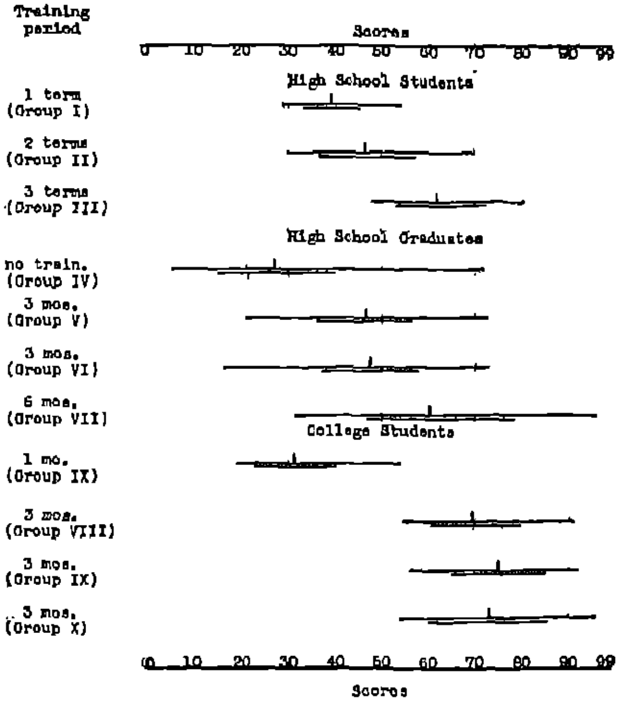


FIG. 2. Levels of radio training, showing M, one σ, and range in scores on Form B for various amounts of radio training classified according to levels of general education.

Analysis of Reliability and Validity

The measurement of progress having been indicated the next step was to check the equality of the forms and verify the reliability at the various levels of specialized radio training, and the different levels of general education; and to establish the validity of the test.

Equality of the Forms. The two forms were compared at the different levels of specialized radio training within the levels of general education. Both forms were found to be equal in difficulty, with variability of performance being much the same for both forms, with the exception of Group I at the general education level of High School Students, performance on Form A being slightly more variable.[1]

Reliability. The odd-even correlations, corrected by the Brown-Spearman formula for length, for all groups tested are given in Table 4. The parallel form correlations for those groups which were tested with both forms is given in the same table.

Table 4

Odd-Even and Parallel Form Correlation at All Levels of General Education

	Odd-Even Correlations							
	Form A				Form B			
Training Period	Raw r_{oe}	P.E.	r_{oe}*	P.E.	Raw r_{oe}	P.E.	r_{oe}*	P.E.
Group I	.53	.00	.69	.07	.53	.09	.69	.07
Group II	.82	.04	.90	.02	.51	.09	.07	.07
Group III	.80	.04	.88	.03	.88	.03	.93	.02
Group IV					.80	.02	.89	.01
Group V	.70	.03	.83	.02	.69	.03	.81	.02
Group VI	.63	.03	.78	.02	.07	.03	.80	.02
Group VII	.84	.02	.91	.01	.82	.02	.90	.01
Group IX (1 month)					.74	.04	.85	.03
Group IX (2 months)	.54	.06	.70	.05				
Groups VIII, IX, and X	.86	.01	.93	.01	.89	.01	.94	.01

* Corrected by Brown-Spearman formula.

	Parallel Form Correlations		
Training Period	*N*	r	P.E.
Group I	28	.39	.11
Group II	34	.73	.05
Group III	25	.44	.11
Group VIII	25	.87	.03
Group IX	30	.70	.05
Group X	34	.02	.02
Groups VIII, IX, and X	80	.88	.02

High School Students (Groups I, II, and III): The groups at the lowest level of high school training (Group I) have an odd-even *r* of .69 on both forms. This reliability is unsatisfactory at this lowest level of specialized training and level of general education.

---

[1] In an effort to save space the tables for this analysis have been omitted, but they may be obtained from the author.

The parallel form correlation of .39 for this group is likewise quite low, and is not four times its P.E. Thus any prediction of individual performance on one form as compared to individual performance on the other would likely be no better than chance for members of this group.

The group at the next level of training (Group II) has a reliability coefficient of .90 that is satisfactory on one form while on the other it is much lower (.67) and unsatisfactory. The parallel form correlation is, however, high enough to be acceptable.

The group at the highest level of radio training among high school students (Group III) has reliability coefficients that are acceptable on both forms of the test. However, the parallel form correlation of .44 is quite low and just four times its P.E.

High School Graduates (Groups IV, V, VI, and VII): The reliability coefficients on both forms for all of the groups at the various levels of specialized training indicate that the reliability of the two forms is generally satisfactory at this level of general education.

College Students (Group VIII, IX, and X): The groups with three months' training were combined on one Pearson Product-moment chart for purposes of computing the odd-even correlation because it was shown that their training was apparently identical as indicated by the means of all groups on both forms.

The odd-even correlation (Form A) at the two months' level of specialized training is lower than for the other levels, with the reliability coefficient for three months' training being most satisfactory. The parallel form correlations for the various groups at the three months' level of specialized training are quite satisfactory and indicate that prediction of individual, and group performance from one form to the other would be highly reliable.

The reliability as shown by the odd-even correlation is acceptable at all levels of specialized radio training, with the exception of the first term students (Group I) at the high school level of general education. The parallel form correlation at the general education level of High School Students is not as reliable as at the general education level of College Students where the coefficients are acceptable for group and individual prediction.

Validity. The progress made in specialized radio training was analyzed at three levels of general education, and within those levels groups were differentiated one from the other by the means according to progress in specialized radio training. Those differences between the means for the various levels of specialized radio training within each general educational level was shown to be statistically significant, and was statistically insignificant between groups with specialized radio training. This distinguishing between levels of specialized radio training indicates a valid test of radio achievement.

Another indication of validity is to be found in the low correlations which were found to exist between the radio achievement test and the Otis Intelligence test. Groups VIII and X were administered the Otis Intelligence Test, Higher Examination Form B, timed for thirty minutes and scored for the number right, and the results correlated with both forms of the radio achievement test. Group VIII has a correlation of −.03 between Form A of the radio achievement test and the Otis Intelligence Test and .26 between Form B and the Otis Test; and for Group X a correlation of .05 was obtained between Form A and the Otis Test and .02 between Form B and the Otis Test. Thus we have indicated by the low correlations, approaching zero in three instances, that there is being measured by the radio achievement test some capacity that is different from that defined capacity measured by this standardized and valid intelligence test.

There being no known acceptable outside criteria with which this test may be correlated, the evidence of parallel form correlation is offered in its stead. Each form of the radio achievement test at the general education level of College Students has a high self correlation, and each form correlates high with the other form. Both forms of the radio achievement test may be said to be valid radio achievement tests and they are measuring equally well that capacity which, for this study, is known as radio knowledge.

Summary and Conclusions

The purpose of this investigation was to measure training progress of radio learners, and to standardize the instrument thus used.

Three levels of general education were represented in the subjects selected for measurement of radio training: High School Students, High School Graduates, and College Students. The subjects at the two highest levels of general education were selected from male applicants on the basis of a Civil Service Aptitude test and a physical examination by Army doctors.

A preliminary standardization of two forms of a radio achievement test was made; the two forms compared for difficulty and variability of performance; odd-even correlation computed for reliability, and a bi-serial correlation of test items computed for the two forms. On the basis of this work the test was revised and the forms reevaluated. Both forms were found to be close to equality of difficulty and variability of performance.

The two forms of the test were administered to each member of three groups, with one, two, and three terms of radio training, at the general educational level of High School Students. At the general educational level of High School Graduates one form (Form B) was administered to a group of novices, and both forms were administered in alternate fashion to the members of the other groups. At the general educational level of College Students the two forms of the test were administered to all members of the three groups; while one group (Group IX) was tested with Form B at the end of one month, and Form A at the end of two months of radio training.

Answers to the following questions were sought from the data thus accumulated:

1. Are both forms of the test reliable at various levels of radio training within the several levels of general education?
2. Is this a valid test of radio achievement?
3. Are the two forms equal in difficulty and variability of performance for various levels of specalized radio training within each general educational level?
4. What prediction of an individual's or a group's progress in training exists for either form?

The answers to these questions are the conclusions derived from this study.

1. The test is a reliable test as is indicated by the odd-even and parallel form correlations shown in Table 4. The odd-even reliability coefficients range from .69 to .93 on Form A and from .67 to .94 for Form B at the various levels of specialized radio training with a median coefficient of .88 for Form A and .85 for Form B. The lowest reliability coefficients were at two levels of specialized radio training within the general educational level of High School Students where the odd-even correlation was .69 for the lowest group (Group I) on Form A and .67 for the middle group (Group II) on Form B. The parallel form correlation for Group I was .39, and .44 for Group III. It appears from this that the test is less reliable at the lowest of the three levels of general education.

2. The test appears to be a valid test of radio achievement at various levels of specialized radio training and general education. The magnitude of the parallel form correlations suggests this as shown in Table 4. The low correlations between both forms of the test and an intelligence test is evidence that the test is measuring some factor other than intelligence, which is assumed to be radio knowledge as that was what the students were studying, and that was what was tested. Essentially however, the test is valid because it differentiates between levels of specialized radio training within each of three levels of general education, and with differences between the mean levels of specialized training that are practically without exception statistically significant as shown in Tables 1, 2, and 3.

3. The two forms are close to equality of difficulty at all the higher levels of specialized radio training for all general educational levels, with Form B slightly more difficult at the lowest levels of specialized training.

4. The coefficients of correlation between levels of specialized radio training at the general educational level of College Students (Group IX) were of such magnitude (.62 to .82) as to be acceptable for group predictions of training progress; and individual prediction as indicated by these same coefficients is increased from 20% to 40% above chance at the various levels of specialized radio training.

In general, it appears that specialized radio training is affected somewhat by the level of general education (see Figures 1 and 2).

Thus it may be concluded that training progress of radio trainees has been measured and a test has been standardized to accomplish it.

Received August 7, 1944.

A Study of the Effect of Music Distraction on Reading Efficiency

Mack T. Henderson, Anne Crews, and Joan Barlow

Grinnell College, Iowa

Very often college students claim that they can study effectively with the radio on, that music does not "bother" them. This suggested the following study which attempts to determine whether or not reading efficiency is influenced when music is used as distraction, and whether there is any difference in the influence of popular and classical music upon reading efficiency. In a similar study Paul Fendrick[1] found that semi-classical music tended to reduce efficiency, but since he did not equate his groups and because he used only one type of distraction, it was decided to supplement his results through this experiment.

Procedure

Fifty freshman women helped us with this experiment. These women were divided into three equally matched groups on the basis of their psychological examination scores (American Council on Education Psychological Examination, 1942 edition) and reading test scores (Nelson-Denny Reading Test) obtained in September, 1943. The A.C.E. means for the No Distraction, Classical, and Popular groups were 107.7 ± 22.6, 102.5 ± 30.3, 103.6 ± 23.8. The means for the vocabulary section of the reading test were 43.1 ± 16.1, 42.6 ± 16.7, 43.8 ± 13.5, and for the paragraph section of the reading test 45.3 ± 11.5, 47.3 ± 11.7, 48.0 ± 13.9. The differences in these means were found to be statistically insignificant as measured by Fisher *t*.

First of all, the subjects filled out a questionnaire which was constructed primarily for the purpose of determining whether or not the subjects were accustomed to studying with the radio on; whether or not they thought that the radio reduced their study efficiency; the amount of studying done with the radio on; and the type of program they usually listened to when studying.

The Nelson-Denny Reading Test was used in this study to measure the reading efficiency of the three groups. This test was chosen for four reasons: (1) it has two forms which made possible the use of one form as a pre-test and the other as a final test; (2) Form A had been given to the freshmen when they entered college, September, 1943, making pre-test scores immediately available; (3) the test has two sections, a vocabulary section and a paragraph comprehension section, making it possible to determine the influence of distraction upon these subdivisions as well as upon the total scores; (4) the test required only 30 minutes to complete, 10 minutes for vocabulary and 20 for paragraph comprehension.

---

[1] Fendrick, P. The influence of music distraction upon reading efficiency. *J. educ. Res.*, 1937, 31, 264–271.

Form B of the Nelson-Denny Reading Test was administered, as the final test, to a group of 14 freshman women with popular music as distraction and to a group of 17 with classical music as distraction, while a third group of 19, a control group, took the reading examination without any distraction. Hereafter, these groups will be referred to as Popular, Classical, and No Distraction groups. Popular and classical music were the two types of music chosen for this study, because the questionnaire results showed that they were the two types to which most of the subjects usually listened. Typical, familiar recordings of both types of music were carefully selected to be played during the tests. The recordings used in this experiment are as follows:

Musical Recordings: Popular music (order of presentation): 1. Two O'clock Jump (Harry James); 2. That's What You Think (Krupa); 3. Sunday, Monday, or Always (Frank Sinatra); 4. Mr. Five by Five (Harry James); 5. Prince Charming (Harry James); 6. Tuxedo Junction (Glenn Miller); 7. Idaho (Benny Goodman); 8. Crosstown (Glenn Miller); and 9. Close to You (Frank Sinatra), and *Classical music:* Symphony in D Minor by César Franck (Philadelphia Symphony Orchestra, Victor Recording, 6726–6730).

The conditions under which the tests were given were regulated as carefully as possible. The tests were administered on three successive afternoons at hours when the greatest number of the subjects would be free. However, it was impossible to find times when they were all free; hence, the rather small number of subjects. During the test, the subjects were asked to assume that they were in their own rooms studying with the radio on. The volume of the phonograph was predetermined by a group of judges, including students, and regulated to approximately the same loudness which the subjects ordinarily maintained when studying with the radio as background. These judges, who were stationed at various places in the room in which the tests were to be given, agreed upon the loudness desired. In order to assure some measure of similarity of volume, the position of the volume control was noted and used throughout the experiment. This method was resorted to, because a physical device for determining volume in decibels was not practical, since the volume within each record varied noticeably.

The significance of the differences in the means was determined by the Fisher t test.

Results

In Table 1 are recorded the averages of the No Distraction, Classical, and Popular groups, the differences between the averages of the pre-test scores and the final test scores of each group, and the significance of these differences in averages. It will be observed that the only score influenced by the distraction more than could be accounted for by chance is the paragraph score of the Popular group. This score was reduced 25.1 score points, on the average, below the pre-test score. It is interesting to note that the vocabulary scores of all three groups showed an increase even though the increases are not statistically significant as measured by Fisher t.

Table 1

Nelson-Denny Averages and t Scores

	N	Pre-Test (Form A)	Final Test (Form B)	Difference	t (Fisher)	P
No Distraction	19					
vocabulary		43.1	50.0	+6.9	1.260	.20
paragraph		45.3	49.2	+3.9	.923	.35
Classical	17					
vocabulary		42.6	48.4	+5.8	.906	.35
paragraph		47.3	46.1	−1.2	.266	.80
Popular	14					
vocabulary		43.8	47.8	+4.0	.605	.55
paragraph		48.0	22.9	−25.1	6.160	<.001

In order to determine whether or not differences exist between those who are accustomed to studying with the radio and those who are unaccustomed to studying with the radio, the data of Tables 2 and 3 are presented. These data show that, regardless of the students' study habits, the groups function alike. The paragraph scores of the Popular group showed a significant decrease in the final test score whether students were accustomed to studying with the radio or not; all other test score changes were within the range expected by chance.

In trying to account for the lack of influence or distraction of classical music upon the test results and the lack of influence of popular music upon the vocabulary scores, one can only suggest explanations. A reasonable explanation for the lack of distraction of classical music is that the rhythms and melodies of classical music are usually more complex and less obvious

Table 2

Nelson-Denny Averages and *t* Scores of Those Who Use the Radio When Studying

	N	Pre-Test (Form A)	Final Test (Form B)	Difference	*t* (Fisher)	P
No Distraction	14					
vocabulary		42.3	49.8	+7.5	1.089	.30
paragraph		44.1	48.9	+4.8	1.062	.30
Classical	9					
vocabulary		41.4	50.4	+9.0	1.341	.20
paragraph		47.8	46.9	−.9	.134	.85
Popular	8					
vocabulary		45.5	50.5	+5.0	.670	.50
paragraph		53.8	25.1	−28.6	5.485	<.001

Table 3

Nelson-Denny Averages and *t* Scores of Those Who Do Not Use the Radio When Studying

	N	Pre-Test (Form A)	Final Test (Form B)	Difference	*t* (Fisher)	P
No Distraction	5					
vocabulary		36.8	40.6	+3.8	.531	.60
paragraph		39.6	40.0	+.4	.042	.95
Classical	8					
vocabulary		43.3	46.1	+2.8	.245	.80
paragraph		46.8	45.3	−1.5	.201	.85
Popular	6					
vocabulary		41.5	44.2	+2.7	.244	.80
paragraph		40.3	20.0	−10.7	3.849	<.001

than those of popular music. The simpler and obvious rhythms and melodies of popular music are easily grasped by a group of subjects and are therefore listened to by the subjects. Naturally, while they listen to the music their attention is diverted from the task at hand. Classical music with its subtle rhythms and hidden melodies is apt to be vague and is therefore not "listened to." It becomes a background against which the assigned task is accomplished without interference, and under these conditions it does not divert the subject's attention from his work. Just what a group of persons highly trained in the understanding of classical music would do under the conditions of this experiment remains to be determined.

A likely explanation for the fact that popular music influences the paragraph scores and not the vocabulary scores seems to lie in the nature of the test materials. The paragraph materials are meaningfully related and require sustained effort on the part of the subject. In contrast to this, the vocabulary materials are intermittent and unrelated. This suggests that popular music interfered with the more complex of the two test sections.

The suggested explanations might be summarized by saying that whether or not music is a real distraction depends upon the complexity of the music and upon the complexity of the test materials. In this experiment, the subtler music (classical) did not influence the test results, and the obvious music (popular) influenced only the paragraph section of the test.

Conclusions

1. Popular music distracted a group of subjects significantly on the paragraph section of the Nelson-Denny Reading Test. Classical music showed no evidence of distraction in either the vocabulary or paragraph sections of the test, nor did the popular music show evidence of distraction upon vocabulary.

2. Students accustomed to studying with the radio were influenced as much or as little as students unaccustomed to studying with the radio.

3. It is suggested that whether or not music serves as a distraction depends upon the complexity of the music and upon the complexity of the test materials.

Received June 2, 1944.

Book Reviews

Note: The length of a review of a book or monograph is no indication of its importance. Because of W.P.B. restrictions on paper reviewers of books have been requested to prepare their reviews in fewest possible words. Ed.

Halsey, George D. *Making and using industrial service ratings.* New York: Harper & Bros., 1944. Pp. xxii + 149. $2.50.

"This book is intended primarily for operating and personnel executives who are interested in the practical aspects of service ratings—what has been done, what the difficulties are, and how rating problems similar to theirs have been worked out successfully by other executives." The book, however, contains material of value to anyone interested in employee ratings.

Ratings of nonsupervisory, supervisory and executive personnel are discussed. Numerous helpful excerpts and sample forms are presented, mostly from the field of public personnel administration.

The author gives proper emphasis to such practical points as how to train supervisors to rate employees and to discuss the ratings with the individuals.

"Service ratings have two distinct and different purposes: to serve as an aid in training and supervision, and to furnish an evaluation of the person's job performance as an aid in making sound administrative decisions—salary increases, layoffs, etc." In developing a rating plan it is important to distinguish carefully between these two purposes. The rating procedure will differ in many respects depending on whether it is designed to serve one or both of these purposes.

The efficiency rating system of the federal government is explained in considerable detail. It is evident that this plan is superior in every way to the graphic rating scale used almost universally in industry.

With regard to the problem of scoring a rating form, the approach is both laborious and naive. From an attempt to attach numerical grades to employee ratings problems arise which cannot be overcome in a practical situation. Logical and statistical considerations, which the author completely overlooks, prevent one from obtaining a total score "by adding the numerical values of the ratings on the separate qualities." There is a wide-spread misconception on this point among industrial people. It is not clear why it is necessary to grade the ratings in order to serve the training or administrative purposes outlined above.

It is unfortunate that the book does not touch on the problem of using employee ratings in a unionized organization. The broader problem is how can the merit of certain individuals be appraised and rewarded under conditions of collective bargaining.

This book is a worthwhile contribution to the literature on rating industrial employees. Although the approach is not scholarly, most of the content is sound and, above all, it is realistic.

Charles C. Gibbons

Owens-Illinois Glass Co.,
Toledo, Ohio

Cason, Eloise Booker. *Mechanical methods for increasing the speed of reading. An Experimental Study at the Third Grade Level.* New York: Columbia University Contributions to Education, No. 878, 1943. Pp. 80. $1.75.

This monograph is a complete report of an experiment set up to answer a specific question for a specific group of subjects: Does intensive training in correct eye movements improve measurable reading skills to a greater extent than would the use of the same amount of time for free library reading? The experiment is admirably designed to produce an unequivocal answer. Paired groups of third grade children (25 pairs in School A, 26 pairs in School B) were used as subjects. Pairing of individuals was based on scores from the Gates Reading Survey Test, but Tables 1 and 2 indicate that the groups were closely similar on the Otis Intelligence Test as well. Twenty-minute training periods were scheduled five times a week for four weeks under the direction of the regular teachers, no other reading instruction being given during the experiment. In School A, experimental subjects read printed material marked and spaced to call attention to phrases. Control subjects read library materials in the same room. In School B, experimental subjects practiced with the Metron-O-Scope. Control subjects read library materials in another room. Before and after the four-week training period, the Gates Reading Survey Tests and several new tests specially constructed to measure skill in the reading of phrases were given. Results bearing on the main problem were definite and clear-cut. *In neither school was there a statistically significant difference on any test between the means of the experimental and control groups.*

The data were then analyzed in several other ways. In order to throw light on the question as to whether the effects of method may differ for children of different ability levels, each group was divided into upper, middle, and lower thirds. Since this would place only about eight or nine persons in each subgroup one would anticipate difficulty in demonstrating that any differences in gains were statistically significant. Therefore, it is interesting to find, using the t-test, that a few were. In the School A Phrase Material Group, the upper third lost in speed while the middle third gained. In the Metron-O-Scope Group the upper third gained significantly more in comprehension than the lower third. In the School B Library Group, the middle third gained significantly more in comprehension than the lower third. Though results based on so few cases cannot be conclusive, they suggest the desirability of special investigation of this problem. Another sort of analysis used the results of a re-administration of the Gates Test after summer vacation. There were still no reliable differences between equated groups. It seemed to this reviewer that the most interesting difference in the whole study was one which the author declined to interpret because of uncontrolled factors in the two situations, namely the difference between schools in gains on the Gates Test. In School A there were no significant gains for either group. In School B, both experimental and control groups showed gains of one-half to three-fourths of a year of reading age, gains which persisted over the summer vacation period. When one remembers that the practice periods added up to less than seven hours altogether, these figures seem very large. Whatever the uncontrolled factors were which accounted for this result in School B may well be of more practical importance than the factors the experiment was set up to test.

There are a few places where the statistical explanations are a little obscure, but there seem to be no errors in procedure which would cast doubt on the dependability of the results. The findings of the study as a whole will lend support to the arguments of those who hold that mechanical and motor skills are less important aspects of reading than are intellectual and motivational factors. Since free reading is much simpler and less expensive than the mechanical methods, the results should have definite practical application.

Leona E. Tyler

University of Oregon

Sargent, S. Stansfeld. *The basic teachings of the great psychologists.* New York; New Home Library, 1944. Pp. xiv + 346. $.69.

The book covers about all the major topics in psychology with a pretty fair balance. It includes most of the applied field as well as the general field, but the former is usually somewhat implicit in the discussions of the latter rather than being set off by itself as applied psychology. In each chapter the treatment is essentially chronological, comprising a brief summary of the findings or theories of quite a number of people on the topic under consideration. The treatment is not critical, and the controversial questions are left open.

The author indicates that it is written for the layman and secondarily perhaps as a refresher for advanced students. The reviewer believes that these two markets should be reversed in importance. The layman will find the discussion somewhat cluttered by the names of so many people unfamiliar to him, and he will not find an extensive topic in which he might be interested treated at any length. It is, of course, impossible to cover so much ground in any very systematic fashion. Occasionally, too, some technical term is introduced without any explanation, although later in the book it may be discussed in more detail. On the other hand, the advanced student might find it quite helpful as a sort of thumbnail review of the whole field, and he would be almost certain to encounter some bits of information here and there that would be a helpful supplement to his existing knowledge on a topic. The author must have covered a tremendous amount of source material in order to formulate the concise paragraphs which he does on so many topics by so many people. From that standpoint the job is well done.

The title of the book is subject to criticism. There certainly are not as many "great" psychologists as listed. The biographical notes at the end include some two hundred psychologists. It is probably not the author's fault, but the publisher's blurb on the jacket is in poor taste and smacks of "gold brick."

Harold E. Burtt

Ohio State University

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

And now to live again. Betsy Barton. New York: D. Appleton-Century, 1944. Pp. 150. $1.75.

Psychology of personnel. Henry Beaumont. New York 3: Longmans, Green & Co., Inc., 1945. Pp. 300. $2.75.

Employment opportunities in characteristic industrial occupations of women. Elizabeth Benham. Bulletin No. 201, Women's Bureau, U. S. Dept. of Labor, 1944. Washington 25, D. C.: U. S. Government Printing Office. Pp. 50. $.10.

Your place at the peace table. Edward L. Bernays. Duell, Sloan & Pearce, Inc., Dept. 3F-12, 270 Madison Avenue, New York 16, N. Y. Pp. 60. $1.00.

Psychology: Principles and applications. T. L. Engle. Yonkers-on-Hudson 5: World Book Co., 1945. Pp. 549. $2.12.

The classification of Jewish immigrants and its implications. A survey of opinion. N. Goldberg, J. Lestchinsky, and M. Weinreich. Yiddish Scientific Institute—YIVO, 535 West 123rd St., New York 27, N. Y., 1945. Pp. 154. $2.00.

The people's choice. Lazarsfeld, Berelson, and Gaudet. New York 16: Duell, Sloan & Pearce, Inc. $3.00.

Braille and talking book reading: A comparative study. Berthold Lowenfeld. American Foundation for the Blind, Inc., 15 W. 16th St., New York 11, N. Y., 1945. Pp. 53. $1.00.

Problems of the postwar world. A symposium edited by Thomas C. T. MacCormick. New York 18: McGraw-Hill Book Co., Inc., 1945.

Human nature and enduring peace. Edited by Gardner Murphy. Third Yearbook of the Society for the Psychological Study of Social Issues. New York: Houghton Mifflin Co., 1945. $3.50.

Jobs for the physically handicapped. Louise Neuschutz. New York: Bernard Ackerman, Inc., 1944. Pp. 204. $3.00.

Selling as a postwar career. David R. Osborne. Chicago: The Dartnell Corp., 1944. Pp. 83. $1.00.

Understanding as a condition for success in order-giving. Paul and Faith Pigors. Industrial Relations Associates, Inc., Cambridge, Mass., 1945. Pp. 28.

Soldier to civilian. G. K. Pratt. New York 18: McGraw-Hill Book Co., Inc., 1945. Pp. 233. $2.50.

The New York Hospital: A history of the psychiatric service, 1771–1936. W. L. Russell. New York 27: Columbia Univ. Press, 1945. Pp. xiv + 594. $7.50.

Intelligence and its deviations. Mandel Sherman. New York 10: The Ronald Press Co., 1945. Pp. 300. $3.75.

Guidance and personnel services. Ruth Strang and Robert Hoppock. Occupational Index, Inc., New York University, New York 3, N. Y. Pp. 6. $.25.

Psychology in education. J. B. Stroud. New York 3: Longmans, Green & Co., 1945.

Techniques of guidance. A. E. Traxler. New York 16: Harper & Bros., 1945. Pp. 381. $3.50.

The job of the industrial counselor for women. Frances W. Triggs. Washington, D. C.: U. S. Office of Education, 1944. Pp. 32. Free.

Elementary statistics for students of education and psychology. E. B. Van Ormer and C. O. Williams. New York 3: Longmans, Green & Co., Inc., 1945. Pp. 120. $1.75.

A study of transfer of training from high school subjects to intelligence. A. G. Wesman. New York: Bureau of Publications, Teachers College, Columbia University, 1945. Pp. 82. $1.75. (Contributions to Education No. 909)

Psychology of teaching. A. D. Woodruff. New York 3: Longmans, Green & Co., Inc., 1945. Pp. 200. $1.75.

Counseling for veterans. Proceedings of Conference January 18–19–20, 1945, sponsored by Minneapolis Vocational Guidance Association, Inc., Minneapolis 14: Center for Continuation Study, University of Minnesota. Pp. 151. $1.00.

Experience in retraining on the Dvorak keyboard. American Management Association, 330 W. 42nd st., New York 18, N. Y. $.75. (Supplementary Office Management Series No. 1)

How to prepare and publish an employee manual: Based on a survey of company practices. New York 18: American Management Association. $.75.

The negro worker: An analysis of management experience and opinion on the employment and integration of the negro in industry. New York 18: American Management Association. $.75.

Supervision of women on production jobs: A study of management's problems and practices in handling female personnel. New York 18: American Management Association. $.75.

Journal of Applied Psychology

Vol. 29, No. 5 October, 1945

The Measurement of Supervisory Quality in Industry *

Quentin W. File

Purdue University

In any organization engaged in a productive activity there must be individuals who assume responsibility for directing productive procedures at the operational level. The Army has its sergeants; the school, its classroom teachers; and the church, its ministers and priests. Industry, likewise, must rely on its all-important directors of operations, its supervisors. Upon these supervisors rests the responsibility of seeing that all the creative planning, technical research, and personnel policies bear tangible fruit in the form of maintained or improved productivity. Operating supervision is the connecting link between top management and the worker, and the chain of harmonious industrial relations can be no stronger than this key link.

In order to establish a common basis of understanding, let us define an industrial supervisor as an individual who actually directs the productive processes at the scene of operation. Such a definition would include individuals between the group leader and departmental supervisor levels.

To meet the wartime demand, supervisory training programs have been instituted both by industry and by the government. Most of the training has been carried out on an "on-the-job" basis so as to provide a minimum of interference with regular work activities. In spite of the extensive publications describing the various types of training programs and testifying to the merits of each, little has been said about the need for objectively evaluating the outcomes of such programs or about setting up a systematic method for evaluating supervisory quality.

---

* This article is based on the author's thesis of the same title, submitted to the Faculty of Purdue University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, June, 1944. This study was carried out under the direction of Professor H. H. Remmers in collaboration with ten industrial concerns. Funds for this research were provided by the Division of Education and Applied Psychology of Purdue University.

The need for some instrument for measuring supervisory quality becomes apparent when one considers the uses to which such a test could be put. A good supervisory ability test could be used: (1) to select and classify candidates for supervisory training; (2) to evaluate the outcomes of supervisory training; (3) for upgrading; (4) to check on the quality of present supervisory personnel; (5) as a basis for interviewing and counseling supervisors; and (6) as material for group discussion at supervisory meetings.

Management was long prone to assume that to obtain the best supervisor one should promote his best worker. In other words, it assumed success on a given job to be a reliable measure of the ability of the individual to supervise others who do the same kind of work. That such is not the case has been repeatedly proved. Ability to deal with personalities, loyalty to management, interest in work and numerous other factors make it necessary to consider knowledge of a job and other specific factors merely an essential part rather than all of the requirements for good supervision.

The general factors of supervision are many. Each supervisor, regardless of rank and experience, must deal with the attitudes of his workers, his associates, and his bosses. He must administer company policies. He must decide what explanations of wage differentials, vacation preferences, penalties, and the like must be given. And, most of all, he must be sensitive to potential disunity and dissatisfaction among his workers in order that the difficulty can be corrected before it reaches a point where production will be affected.

Construction of the Test

In the construction of an instrument for measuring supervisory quality, consideration must be given to the relative importance of both the general and the specific factors involved. The seemingly predominant importance of the general factors of supervision is emphasized by the large number of books and articles now expressing the need for improved understanding of human relations and the importance of personalities in achieving industrial harmony. The following statement by Dodd and Rice[1] is illustrative of present personnel trends: "When it becomes necessary, new supervisors are selected from the ranks of workers, engineers, technicians, and other sources. Experience proves conclusively that intelligence, personality, vitality, and leadership should outweigh technical or trade ability when the selections are made."

Most industrial supervisors are obtained by some form of upgrading. It seems quite possible, therefore, that any individual, who is able to

[1] Dodd, A. E., and Rice, H. O., eds., *How to train workers for war industries*, p. 80.

qualify for a supervisory position on the basis of his general abilities, will either have acquired, or can acquire, the specific knowledge necessary for handling the job. It was on the basis of the hypothesis that *factors generally common to industrial supervisory positions are the really important quantities* that this project of constructing a valid measure of supervisory quality was conceived.

When developing any test, it is necessary to make certain basic assumptions. The principal assumptions of this study were:

1. That ability to supervise workers is something general in nature rather than highly specific to a given job or company. The supervisor's effectiveness is, in the long run, dependent upon his understanding of and ability to deal with *human relations.*

2. That lack of this general ability to deal with workers is the greatest single cause of supervisory failures and of management-worker friction.

3. That knowledge of how to handle the supervisory function can be tested by obtaining responses to certain significant questions which are drawn directly from problems which frequently confront the supervisor.

4. That such questions can be obtained by direct contact with supervisors on the job, by careful study of the literature concerning supervisory fundamentals and supervisory problems, by taking into account the relevant principles of psychology, and by systematically "weeding out" those items which prove unfruitful.

In selecting the items for the supervisory ability test, *How Supervise?*,[2] three definite objectives were kept in mind.

1. The items must be presented in problem form calling for an operational response, i.e., the items should ask "What should be done . . . ?" or "Is it desirable to . . . ?", etc.

2. The items must have "face" as well as statistical validity. They must present problems which are pertinent to industrial supervisors regardless of the department or the company from which the supervisors are selected.

3. These items must be simply worded so that *any* supervisor can see the problem involved.

Item Selection

Items for *How Supervise?* were selected from three distinct sources: publications concerning industrial supervision, suggestions from industrial supervisors and personnel men, and contacts with labor leaders. The most fruitful and readily available source of potential items was the industrial literature. Industrial supervisory problems have received

[2] Sample copies of this test may be obtained from the Psychological Corporation, 522 Fifth Ave., New York City.

considerable attention in the last half decade and much has been written about the importance of harmonious supervisor-worker relations. Especially valuable of the published works were the human relations manuals and books dealing with specific supervisory problems and their solutions.[3] Contacts with various supervisors in a sizable manufacturing concern offered a means of checking on the practicality of the problems presented and of obtaining additional items.

From the sources mentioned above a pool of 204 items was gathered for the experimental edition of *How Supervise?*. These items were divided into two numerically equal forms. Where items dealt with closely related problems, those items were placed in different forms. All other items were divided on an odd-even basis. Many of the items deal with problems now under heated discussion. These items were purposely included in order to give the test added value as a basis for supervisory conferences.

The personal data which the supervisor was asked to provide were purposely made extensive in order to investigate possible relationships of these data with test scores and with management's ratings. This section was purposely placed at the end of the test for three reasons.

1. The supervisor will be more likely to give his undivided attention to oral instructions at the beginning of the test if not presented with preliminary material to be filled out before answering the test items.
2. Any resentment which might occur from a fairly extensive set of questions will not affect his responses to the items of the test.
3. Since tests are usually scored from front to back, the total score and the name of the supervisor are brought together with a minimum of page turning. This advantage was also evident when responses to the items were punched on tabulating machine cards.

Item Validity

The validity of a test item must always be described as validity with respect to some standard of value. One of the most vital, and usually the most difficult, problems of test construction is that of securing an adequate criterion. One criterion for a test of supervisory quality is obviously "success on the job." Success, however, must be defined in terms of some standard and by some individual or group of individuals.

The problem also arises as to what are the best answers to the items of a test. To meet the above problem two assumptions were made:

1. "Good" supervisors *as a group* know the best answers.

---

[3] Gardiner, Glenn L., *Better Foremanship*, First Edition. New York: McGraw Hill Book Co., Inc., 1936. Heyel, Carl, *Human Relations Manual for Executives*, New York: McGraw Hill Book Co., Inc., 1939.

2. Men who write books and articles on industrial supervision and men actually engaged in directing supervisory training programs *as a group* know the correct answers.

Unfortunately, simply to assume that the best supervisors know the correct answers, though logically sound, does not provide an adequate criterion. The really crucial problem becomes one of determining who these good supervisors are. The writer felt that two groups of people should know—(1) the men who work under the supervisor and (2) the individuals to whom the supervisor is responsible: in other words, his bosses.

Ratings by members of management above the supervisor proved more available than those of the workers. Out of a total of 972 supervisors tested 577 sets of four ratings were obtained. The instrument used for rating was *The Purdue Rating Scale for Supervisors* constructed by H. H. Remmers and the writer for use on this study. This scale asked for an evaluation of the supervisor in terms of seven factors and an overall evaluation of his quality when all factors were considered.

The orthodox method of test validation consists of having so-called "experts" answer the items of the test and using their responses as a key for obtaining a total score for all individuals tested. This total score is then used as a criterion for determining the degree of discrimination possessed by each item.

Foremost of the problems in obtaining expert judgments is the problem of determining who the experts are. Two groups of individuals were sampled to get the expert judgments needed to provide a scoring key. Group One of the sample of experts consisted of eight individuals who had either written articles or books about industrial problems, or were recognized authorities in the field of mental hygiene. To insure careful consideration of the test, a check for $10 was enclosed with each set of materials sent out.

Group Two of the experts consisted of thirty-seven individuals working for the government under the Division of Vocational Training for War Production Workers. Twenty-four different states were represented with no more than five individuals from any state. No financial reimbursement was given any member of this group and all responses were on a purely voluntary basis. Both groups of experts were asked to criticise each item and offer suggestions for improvement as well as to give the answers they considered best.

All scoring of the supervisors' responses was done from tabulator cards. A scoring key was obtained by finding the responses most frequently judged best by the two groups of experts combined. Combining the responses of the paid experts with those of the Training

Within Industry group was justified on the basis of a correlation of + .91 ± .01 between the modal responses of each group. Items about which the experts were unable to agree and those on which the modal responses were "uncertain" were not used in obtaining the total test score for the supervisors.

The average supervisor tested for this study was thirty-four years old, married, with one or two dependents, and probably didn't own his home. He had a high school education and had taken a supervisory training course. He had worked nearly ten years before becoming a supervisor and over six years since becoming one. He was in charge of forty-nine workers of both sexes. In the last ten years he had been employed by two different companies.

Further analysis revealed that: only 2% of his fellow supervisors were women; 15% of all supervisors had worked less than two years before being promoted and that 38% had been supervisors less than two years; only 5% were single while almost 20% had no dependents; 45% owned or were buying their homes; 23% had some college training and two-thirds of the group had completed supervisory training courses; about 30% had been with their present employer for at least ten years or else had never worked anywhere else; and 73% supervised both men and women while only 2% supervised women alone.

These supervisors were drawn from ten industrial concerns. The number of supervisors contributed by each concern was a function of the size and organization of that company.

Tests of the discriminating power of each item of the experimental edition of *How Supervise?* were made with respect to both of the criteria previously described, namely, managements' ratings and total scores on the test. The method used to determine each item's discriminating power was the critical ratio of the difference between the average responses of the upper 27% and the lower 27% of the supervisors with respect to the criterion. This method was favored because it involved no assumptions as to the right and wrong answers to individual items. To the extent that a given item yields significant differences with respect to an acceptable criterion, the item can be considered valid and its correct answer will be indicated by the direction of the difference between the upper and lower groups. Thus, as was the case with this study, when a fallible criterion is used, the validity coefficient obtained for a given item depends only on the item's value and the general validity of all of the experts' judgments. Weakness of the experts' responses to any one item does not make it impossible for the item to show a significant discrimination ratio.

When selecting items to be retained for the final forms of *How Supervise?* four different factors were considered:

1. Size of the critical ratios of the differences between the upper and lower groups with respect to total score on the test.
2. Size of the critical ratios with respect to management's ratings.
3. Degree of discrimination throughout the continuum of possible answers, i.e., difficulty of the item.
4. Company to company variation on the item.

Most important of the above factors used to determine the value of the items were the critical ratios of the differences with respect to total score on *How Supervise?* This criterion was favored because of the high agreement between different groups of experts and significant differences between the upper and lower groups of supervisors.

Study of the Criterion

In this study considerable time and statistical attention was given to management's ratings. It was hoped that a criterion could be obtained which would be independent of the items of the test itself. The desirability of securing this independent measure of supervisory quality is readily apparent when one reviews the advantages which such a criterion offers.

1. Management ratings are made at the actual scene of operation. In addition to being independent of the experimental edition of the test, such ratings should constitute a measure of actual success on the job.
2. Since management is solely responsible for determining what individuals will be promoted to supervisory positions, much can be said for selecting those individuals who measure up to management's standards.
3. Management as a group should be more adept at making ratings since most modern industrial organizations make use of some form of merit rating.

The 577 supervisors rated for this study were employees of six different industrial concerns, ranging in size from 500 workers to 20,000 workers. Reliability coefficients and intercorrelations of the rating scale items were computed for the entire population tested. In general, the correlation between the traits of the rating scale tended to be greater than the reliabilities of the traits correlated. At least 17 of the 28 item intercorrelations significantly exceeded 1.00 when corrected for the unreliability of the items. Since no correlation above 1.00 can exist in practice, it must be assumed that corrections for attenuation are not applicable to these data. The formula, $r_{x\infty y\infty} = \frac{r_{xy}}{\sqrt{r_{xX} r_{yY}}}$, corrects for chance unreliability of the items, but, like all other formulae assuming random sampling, makes no allowance for constant or systematic errors. *This formula is*, therefore, *not applicable to data in which these constant errors occur.*

Many authors of statistical texts have assumed that there can be no significant correlation between completely unreliable measures of a given pair of traits.[4] When corrected correlations above 1.00 were found, these correlations were assumed to be due to sampling error. It was assumed that further sampling would reveal either lower item intercorrelations, higher item reliabilities, or both.

The writer advances the hypothesis that under conditions where excessive halo[5] and logical errors exist, these spuriously high intercorrelations between traits can be obtained from successive samples of randomly selected ratings. The only requirement for such a condition is that the logical relation of one trait to another be greater for all raters than the reliability of those raters' judgments.

Four tests of refined methods of scoring the Management Ratings were made. In the data tested all attempts at statistical refinement failed to reveal significant increases in the reliability of the ratings. The methods tested include weighting by Beta weights, z-scores, and item reliabilities.

Of the 204 items of both forms of the experimental edition of *How Supervise?* 23 items yielded critical ratios of 2.00 or better for the differences between the upper and lower groups as selected by management ratings. Since less than five items could be expected to occur by chance under these conditions, some idea of what management expects of its supervisors can possibly be gained by examining the nature of these items.

Assuming that the item preferences of management-selected supervisors do reflect management's opinions of what constitutes good supervision, the following observations can be made. Management believes:

1. Its supervisors should accept responsibility for keeping their department's production up and its costs down.
2. In standardized procedures even to the point of failing to recognize the importance of individual differences among workers.
3. In deeds rather than words.
4. Its supervisors' first and foremost responsibility is to management.
5. Fines are not the best way to discipline workers.
6. Workers should not be given regular rest periods.
7. Procedures for granting raises and promotions are management's business, not the workers'.

---

[4] Peters, Charles C., and VanVoorhis, Walter R., *Statistical procedures and their mathematical basis.*

[5] Tiffin, Joseph, *Industrial psychology.*

Probably most significant of the findings on the management-rating criterion is the absence of any significant critical ratios on items dealing with the mental hygiene aspects of supervisor-worker relations. Most of these items were highly significant on the total score criterion. One may well wonder to what extent the principles of improved personnel relations have trickled through to operational levels.

To investigate the relation of management's ratings to personal information about the supervisors, the correlations between ratings and age, marital and home status, education, working experience, special training, number of men supervised, and number of companies worked for were computed. Only three of these correlations were significantly above zero at or above the 5% level of confidence. There appears to be a slight tendency for supervisors, with a great deal of working experience before being promoted, to receive somewhat higher ratings. Supervisors who have taken supervisory training courses are rated slightly higher. More significant, however, is the relation between number of men supervised (an indication of rank) and management's ratings. This relation is a natural one since it is generally assumed that promotions are most frequently given to individuals whom management considered best.

Validity of Management's Ratings

Since the analysis of results on *The Purdue Rating Scale for Supervisors* revealed management's ratings to be of questionable validity, *these were used as a secondary criterion for the validation of the test* items rather than as a principal basis for the selection of discriminating items. Reasons for skepticism as to the validity of this criterion are listed below.

1. Unusually large halo effect indicating that only one general factor was being measured.
2. Relatively low reliability of total scores on the ratings. Near-significant increases in reliability were obtained where corrections for differences in judges were possible. Since this correction could not be made on 80% of the ratings, a known source of sizable errors did affect the validity of the criterion.
3. Variations in quality of raters from department to department doubtless existed. Such variations would tend to increase the spread of the rating scores and thus increase the computed reliabilities.
4. Almost complete failure to find significant discrimination values on supervisory problems recognized by industrial experts as important; items which two groups of experts were able to agree upon.

Several hypotheses can be advanced for the failure of management's ratings to prove their worth as a criterion. The following seemed most tenable to this writer:

1. The pyramid of authority inherent in industrial organizations does not provide the necessary contacts for multiple rating by management. Sound organization, according to this hierarchy, requires that each supervisor is responsible to only one individual. Requiring four men to rate a supervisor probably necessitates calling in raters who have had little contact with the person they rate.

2. Rating conditions are more difficult to standardize than testing conditions. Inherent in these ratings are such factors as personal relations between rater and supervisor, personality characteristics of the rater, his experience in rating, and company attitude toward merit rating.

3. Varying standards adopted by raters in a highly skilled department as compared to raters in a non-technical division and company-to-company variations can account for extremely wide variations in rating scores which have no basis in terms of supervisory quality.

4. Quite possibly either or both management rating and *How Supervise?* deal with only part of the required abilities necessary for good supervision. For example, operational management has long disregarded the area of mental hygiene for workers and the effects of employee attitudes. *How Supervise?* deals primarily with this area.

Judgment of Experts

In addition to being asked to provide the best answers to the experimental edition of *How Supervise?*, each of the industrial experts was asked to mark items which he thought were ambiguous or of no value and to indicate any corrections or comments he cared to make. Each expert was also asked if he thought the language used in the items would be understood by supervisors and if the approach to the problem was a practical one. Over 90% of the experts thought the language was sufficiently clear, while between 75% and 80% thought the approach to the problem practical.

Two indications of the reliability of the experts' responses were obtained: (1) a correlation of .91 between the modal responses of the paid experts and those of the Training Within Industry experts, (2) a correlation of .80 was obtained between the experts' scores on Form A and their scores on Form B when scored back on the key derived from their modal responses. This would indicate that the reliability of the total score criterion was approximately .89.

Traditional requirements for test validity were well satisfied in this experimental study of *How Supervise?*. Experts in the field in which the field in which the test is designed to operate were able to agree as to the correct answers to the items. The reliability of the experimental edition was found to be .84 ± .01 for scores on the two forms of the test combined

($N = 577$). Wide variations in the total scores made by supervisors were found ranging from near chance to almost complete agreement with the scoring key.

As would be expected, wide company differences in the average quality of supervisors, as measured by total test scores, were found. Of the forty-five differences between the average test scores made by supervisors of the ten companies tested, 15 were significant at the 10% level, 10 at the 5% level, and 7 at the 1% level of confidence. This would indicate that some factor or factors were measured by the test which exist in varying amounts in different companies. This is especially interesting since significant differences were also found between the scores of given groups of supervisors at the beginning and end of training periods. The significant improvement measured in the latter situations indicated that the test was measuring an improvement which occurred during this period and that there was an overlapping between material covered in the course and the content of *How Supervise?*.

Validity of *How Supervise?*

Briefly summarized, the experimental indications of the validity of *How Supervise?* are:

1. Supervisory achievement in industrial training courses has been measured and significant improvements found.

2. Areas which industrial experts consider vital have been reliably measured with test items about which the experts agree. Coefficient of Reliability = + .84 ± .01.

A study was made of the relation between total score on *How Supervise?* and such personal information as marital status, age, education, number of men supervised, etc. Several correlations were found which were significantly above zero but not of sufficient size to indicate that personal data would be of importance in selecting good supervisors. A correlation of + .35 between education and total test scores is the only relation of sufficient size to be of importance in the selection of supervisors. The optimum amount of correlation which should exist between education and a test of supervisory quality is problematical. While such a test should not correlate highly with amount of education, doubtless, formal education does provide valuable learning situations which are generally helpful.

It is interesting to note that the greater proportion of this correlation resulted from differences between supervisors who had college training and those who did not have college training. For example, 44% of the elementary school graduates were above the 50th percentile on the overall norms, and 50% of the high school graduates were above this point. At the college level, however, 69% of the supervisors who completed one year of college and 74% of the college graduates were above this median score for all supervisors. This would seem to indicate that selection at the college level tends to "weed out" individuals who have failed to develop an understanding of the general factors of human relations, or that colleges in the student's first year of training provide considerable opportunity for gaining insight into human relations problems.

No major changes were made which seemed likely to cause the supervisor to place a different interpretation on the problem presented. Revisions were made only on items whose relation to the total score criterion was significant and which comments by either the experts or the supervisors tested indicated some confusion in interpretation.

Items not included in the final form were disqualified for the following reasons: lack of discriminating power, too easy, or weak on one or more criteria.

Each form of the final edition of *How Supervise?* contains 70 items which are divided into three categories. Care has been taken to make each division of a given form equivalent to its corresponding division in the other form. Each division is equated on the following factors:

1. Variability of the item—standard deviation of all supervisors' responses.

2. Discrimination index—critical ratio of the difference between the mean responses of the upper and lower groups.

3. Difficulty index—deviation of the average response of all supervisors from the correct response. These values were computed on the basis of a three-answer continuum.

4. Number of positive and negative items in each category of each form.

Quite naturally, industrial participation in this experimental program was not of a benevolent nature. To insure that the companies, as well as we, would receive appreciable benefits, the following reports were sent to each cooperating concern.

1. Scores for each supervisor on the test with percentile values based upon norms for all supervisors tested in this study

2. Individual reports on the quality of each supervisor as rated by four members of management.

3. Summaries of supervisors' scores on the experimental edition of *How Supervise?* by departments where a sufficient number of supervisors were tested to make such breakdowns meaningful.

4. An overall summary of the company's scores on the test with an indication of their relative position with respect to other companies tested.

5. An item-by-item tabulation of the per cent of the company's supervisors who gave each of the possible responses to that test item.

6. Where forms of the test were given at the beginning and again at the end of a training period, comparisons between each supervisor's scores were given, together with an overall evaluation of the program as a whole.

Summary and Conclusions

Conclusions drawn from this study can best be made in terms of the hypotheses advanced when plans for the experimental project were conceived. These hypotheses logically fall into three categories: (1) those which deal with the nature of industrial supervision, (2) those which are concerned with criteria against which supervisory quality can be measured, and (3) those which deal with methods of scoring and computing data. Both the hypotheses and findings concerning them are discussed below.

The hypotheses advanced as to the nature of industrial supervision were:

1. *Important aspects of industrial supervisory ability can be measured by test items which are equally applicable to all industrial concerns.* True. 140 discriminating items were found in this study; items which showed no significant variation with respect to the size or nature of the industrial concern. Confidence in the importance of these items was expressed by both industrial experts and management.

2. *The mental-hygiene aspects of industrial supervision are of primary importance.* In other words, *supervisor-worker relations are among the key determinants of good or poor supervision.* True. Several indications of the validity of this hypothesis were found.

a. The average discriminating power of the items of *How Supervise?* which dealt with human relations was significantly greater than the average discriminating power of factual items.

b. In response to a felt need, the last decade has witnessed innumerable publications of books and articles dealing with the human-relations aspects of industrial supervision.

c. Supervisory training courses, which place considerable emphasis on this area, are now being given.

d. The existence of labor troubles, so frequently blamed on conflicting personalities, adds further emphasis to the importance of mental hygiene in industrial relations.

3. *A general test of supervisory ability can be used to evaluate the outcomes of supervisory training programs.* True. The experimental edition of the test was used by two different companies for this purpose. Sig-

nificant gains were found in both cases, especially among the poorer supervisors.

4. *Age, education, and miscellaneous other variables are highly important factors in good supervision.* Generally false. Of all the personal information examined, only education revealed a relationship above bare significance with respect to total scores on the test. It should be pointed out, however, that experience was measured in terms of two-year intervals. Differences which exist between a supervisor of one and a half years of experience and one with no experience at all may well have been overlooked.

The hypotheses advanced concerning criteria for validating the test were:

1. *Four members of management can be found who are sufficiently well acquainted with any particular supervisor to rate his abilities accurately.* Questionable. Ratings obtained for this study were not sufficiently valid for use as a criterion for determining test item discrimination. Differences in standards set by different raters, lack of knowledge about the supervisor rated, and logical error (halo effect) concerning relations between rating traits all tended to make the obtained ratings invalid.

2. *Industrial experts as a group give reliable answers to the problems presented in the test items.* True. Two completely different groups of experts agreed closely as to the best answers to the items of the test ($r = +.91$).

3. *Top management and industrial experts agree on what constitutes good supervision.* False. Validity of this hypothesis would have eliminated the need for two criteria for the validation of test items.

The hypotheses advanced concerning different methods of scoring both rating-scale and test data were:

1. *Weighted scoring of ratings significantly increases the reliability of the total rating scores.* Generally false. The only appreciable increase in reliability, which resulted from the previously described weighting methods, was that of correcting for the variability of individual judges. This increase in reliability was only significant at the 11% level of confidence, and was not applicable to most of the data.

2. *Test items which provide five possible responses to each item yield more reliable measures of supervisory quality than items which provide only three possible responses.* False. Identical reliabilities were obtained for the two types of items. On the basis of this finding, items in the final forms of *How Supervise?* provide for only three possible responses, "agree, uncertain, disagree."

In addition to the hypotheses accepted or rejected, other observations were made for the analysis of the experimental data. Assuming that management-selected supervisors do reflect the attitudes of top management in their responses, the following observations can be made:

1. Management and industrial experts significantly disagree:

a. On methods of handling dissatisfied workers. Industrial experts favor transfer; management opposes.

b. On methods of handling complaints. Management favors standardized procedures for each type of complaint; the experts favor the recognition of individual differences.

c. As to the desirability of delegating responsibility to workers for improving working conditions. Management opposes.

d. As to the wisdom of allowing regular rest periods. Management opposes.

e. As to whether a worker should be told what promotions he can expect providing he attains a certain level of proficiency. Management maintains that these matters of salary and promotion are company business which should not be disclosed.

2. Industrial supervisors, selected by management as best, are not fully aware of the importance of human-relations problems in industrial supervision. Very few of these problems as presented in the test items approached significance with respect to the management-ratings criterion. The same items were highly significant with respect to the total score criterion.

From the hypotheses investigated and observations made, we may conclude that general factors of supervision do exist and that these quantities can be measured. The human-relations aspects of supervision are vital and are, of necessity, receiving an ever-increasing amount of attention from management. Industrial experts, both theoretical and practical, have rather clear-cut ideas about these general factors. Industrial management tends to be less progressive and seems to favor keeping the worker "in his place," rather than encouraging him to become interested in "company affairs." Management's idea of what it wants in a good supervisor seems rather inclined toward negative rather than constructive methods of handling supervisor-worker relations. Management is, however, well aware of the factual problems in industry and how they should be handled. Only on items dealing with the mental-hygiene aspects of supervision were there indications of significant weaknesses.

From this study, a test of the general aspects of supervisory quality has been developed. It is believed that this test, *How Supervise?*, will prove valuable for selecting candidates for and evaluating the outcomes of supervisory training programs, for selecting individuals for direct promotion to supervisory positions, and for checking on the quality of present supervisory personnel.

Received September 23, 1944.

Personnel Placement in the Armed Forces *

John M. Stalnaker

Stanford University, California

It would be presumptuous for me to pretend to a comprehensive knowledge of the many and varied methods now being used in the selection and placement of personnel in the Armed Services. Many groups in both the Army and the Navy have been concerned with these problems. The civilian Office of Scientific Research and Development of the Federal government and certain quasi- and non-governmental civilian agencies have made contributions. It may be permissible at this time, however, to review some of the trends and mention some of the recurrent problems.

It must be recognized that those of us who have been continuously working on the problem of selection of service personnel suffer at this time from lack of perspective. We are too close to the work and therefore apt to magnify what will ultimately seem like minor operating difficulties. The fact that most of the data and tests are at present "classified"[1] further limits what can be said of them. Virtually all of the tests, as well as the research and the investigational studies, have been financed by the government, and all results are therefore subject to strict governmental control. Government employees and others having access to the data are not permitted to report on such results without official permission, and no such permission has been sought for any matters covered in this paper.

The importance of the wise use of manpower has probably been recognized to a greater extent in this war than at any previous time. The examples of waste and extravagant use of manpower which can be cited may seem to deny this statement. The cost-plus type of contract does not encourage economy in the use of civilian manpower, and certain branches of the military and naval departments have at times deemed it wise to reserve or hoard desirable men. The fact still holds, however, that in this war more than ever before attention has been paid to putting the man in the job for which he is best suited, and assigning to special training only such men as could absorb the training in the time allowed.

---

* An address delivered at the Cleveland meeting of the American Statistical Association, September 13, 1944.

[1] The term "classified," as applied to government documents or data, signifies that the material is either "restricted," "confidential," or "secret"—i.e., available only to a small number of specified individuals.

Random selection, which was not unknown in the services—for example, selecting every other man, or the first ten or twenty men on the muster roll—has been replaced by methods clearly superior. Some of the more satisfactory devices are costly both in time and in dollars—for example, assigning any individual at random to try his hand at a complex task and retaining him only after he has clearly demonstrated his ability. Such methods are being replaced by more efficient and economical means of selection which yield demonstrable savings.

The increase in use of appropriate selection devices is a relative matter. There are still to be found able and experienced members of our armed forces who believe that, in selecting the fighting man, only the imponderable personal characteristics are of importance and such variables can quickly be recognized—not measured—by only a few men steeped in the tradition of the particular branch of service concerned. Even such die-hards, however, are gradually coming to recognize that through scientific selection—including test scores, an evaluation of background and training, an estimate of certain personality traits, and a recognition of interests—men are now being trained better and faster. And fast technical training, no one will deny, has been of the essence. The fighting men of today must be technicians who are well trained in the operation and maintenance of complex instruments.

The most striking evidence that the importance of modern methods of selection and placement is being recognized is found in the number of psychologists employed in the task. The work in the office of the Adjutant General, in the Air Surgeon's office, in the Medical Research group in Naval aviation, in the Bureau of Naval Personnel, in the Armed Services Institute, in the projects under the Committee on Service Personnel of the National Research Council and its successor, the Applied Psychology Panel of the National Defense Research Committee, and in the selection and placement of men in the Army and Navy college programs—all these attest the increasing role being played by measurement work in the selection and placement of men. Dollars are so freely spent these days that cost figures are less significant, but should someone have the time and authority to calculate the cost of the technical and developmental work being done by the Army and the Navy in selection procedures, the figure would be most impressive. The savings which the use of the techniques so developed have made possible would be even more staggering. It must never be forgotten that no weapon is better than the man behind it—and that modern weapons make heavy demands on training, and skill, and discretion.

There are inherent difficulties in finding desirable and economical methods for the selection of men destined for success in a war activity.

Seldom can all variables be controlled experimentally. Laboratory set-ups are frequently not feasible. In field checks, a multitude of basically irrelevant factors operate to reduce correlation. The requirements for a given school are frequently shifted for reasons not always evident to the research psychologist. Sudden demands from the Fleet, for example, for more men trained in a particular school may force a lowering of the standards for the men taken into the school if, as is usually true, the supply of manpower is limited in quality as well as quantity. The many difficulties characteristic of field study are augmented by the not infrequent and drastic shifts characteristic of the Army and the Navy. A radio code school unexpectedly closed down may leave unfinished a lengthy and costly experiment. The supply of available men may suddenly be changed, rendering a change in standards both desirable and necessary.

There is seldom a clear-cut and unquestioned final criterion against which to validate selection procedures. The criterion of successful performance under combat conditions of the duty for which the man is trained is seldom obtainable. Furthermore, combat conditions are not stable. The combat situation is not the same for all men even in a given type of work and in the same unit. The best sort of criterion from the combat field is usually a rating or estimate by superior officers; and only in rare cases can such ratings, with all their weaknesses, be obtained. Because this criterion is rarely available and can seldom be obtained in any satisfactory fashion, we must accept intermediate criteria. As a matter of fact, so called intermediate or non-combat criteria are entirely suitable for many tasks. Weeks and months of preparation are necessary for a brief period of actual combat. Many service men must regularly engage in duties removed from combat. These prosaic day-to-day jobs of the technical sergeant, the yeoman, the storekeeper, the radioman, and many others are of great importance in making it possible for us to win battles. The skill and zeal they show in their non-combat duties should not be underestimated.

Most technical jobs in the services are restricted to men who have graduated from formal courses usually called schools. The length of a course varies from a few days to several months. To be rated as a torpedoman in the Navy, for example, a new recruit must, after general preliminary training, go to a school for torpedomen. To obtain the coveted wings of the Naval aviator, one must graduate from the extensive aviation training courses. Any economy which can be effected in selecting men who will do well in these schools is obviously worth while. Most selection procedures are designed to pick men who have a high probability of success in the school concerned. If scores on a valid performance test can be obtained for the criterion, so much the better. School grades may reflect many traits of the instructor as well as the students, and so are frequently less useful. If the skill of the bombardier is measured by his hits on a target in fifty standardized practice runs, an unusually good criterion is available against which to validate the selection and training procedures. Valid performance criteria are only rarely obtainable.

If the schools in certain cases have been somewhat out of date, if they fail to achieve proper motivation, if they eliminate men for reasons of discipline or personality, the problems of selection are complicated. If the schools change and improve as rapidly as possible, as they certainly should, the nature of the tests and scores used in selection must be subjected to constant study and revision. Selection and training cannot be separated; they must be dealt with in most cases as a unit.

The first logical step in evolving a program for selection of men to enter a certain technical course or school is to make an analysis of the school curriculum. This analysis is usually made informally and sometimes intuitively. Tests are then selected which are believed to measure the traits deemed essential for success in the school. In selecting the men to enter the school, the classification officer should evaluate past training and experience along with the test scores. Grades in the school, graduation from the school, and ratings obtained are frequently used as criteria for the validity of the tests. Many different tests—the verbal factor, mathematical aptitude and knowledge, spatial tests, tests of general aptitude for electronics, mechanical aptitude, etc.—have been given, and various ways of using the results are available. Conditions within the schools change, as has been pointed out. Thus there are many difficulties in the way of establishing permanent or final methods of selection, but the immediate gains made through the use of recommended procedures have again and again been demonstrated.

Tests and selection procedures which are effective in selecting men from one type of population will not necessarily be equally effective in picking men from another type of population. If only college graduates have been considered for a particular school, a new selection technique may be necessary when men with only a high school education are made eligible. A selection program suitable for older men with considerable trade experience will not necessarily work equally well with men just out of school. As the nature of the pool of men from whom selection is made changes in its essential characteristics, the testing program must be revalidated. For an accurate and complete interpretation of a test score, it is usually necessary to know something about the characteristics of the population being tested.

An illustration may be taken from the Army-Navy College Qualifying Test. An item which is successful on this test (in the sense of predicting total score on the section) for high school seniors from large urban communities in Iowa and Nebraska will not necessarily be equally successful for seniors from small rural locations in New York state. Consider an item in the section on so called common-sense physics, which supposedly tests for knowledge frequently obtained from other sources than the class room. One item concerned the long distance transmission of electric power. This item was approximately of equal difficulty for the two groups—rural and urban—but much more valid for the group from rural New York state. Even in the antonym type of items, significant differences are found in certain cases—and these items cannot be guessed in advance. Why should *domineering* as the opposite of *servile* be easier and more valid for New York city seniors than those from rural Alabama and Georgia? The determination of the angle between the hands of a clock at 4:10 is more valid for seniors from urban centers in California than those from urban centers in New York.

These populations of high school seniors from different areas of the country and from rural and urban environments are similar in many ways. The shifts in the inductee population with changes in draft regulations and with the exhaustion of certain types of eligible men are much greater, and cannot be ignored in interpreting test results.

There are many techniques now available to improve the efficiency of tests and selection procedures. Some method of item analysis has been widely used by many of the technicians. Through the use of some index such as the biserial correlation coefficient, or some estimate of it, or certain empirical indices, an analysis of the behavior of a population on the individual item can be determined. Non-contributing items can thus be eliminated and efficient items retained. The difficulty of the item for the population can at the same time be determined. The criterion for the analysis usually is the score on the total test, although an acceptable external criterion may be even better if it is available. It is to be hoped that the comprehensive study of the interpretation of item analysis which has been done for the services will eventually be made available for wider distribution.

In measuring the success of a test for selection, the simple or multiple correlation is frequently used, with the test scores on the several measures serving as the "independent" variables, and the school grades or success on some performance measure or ratings by superiors as the criterion being predicted. In those cases where the tests are used to eliminate the potentially unsatisfactory rather than to predict the degree of success, critical or "cutting" scores can be used. In such cases the empirical results are shown in a four-fold table of pass or fail on the test, and success or failure in the school. The assumptions underlying the cutting-score procedures are more simple and direct than are those where the tests are used to predict the entire range from the lowest failure to the top success in the criterion; on the other hand, the cutting-score procedures suffer from all the statistical and practical disadvantages of coarse grouping.

Certain personality scales and psychoneurotic inventories have been used to select men for special attention by the psychiatrist, or to eliminate men from further consideration for certain tasks believed to demand stable types of personality, such as submarine crews or paratroopers or men for certain branches of the intelligence service. In these as in many other cases, the selection cost [2] is an item of importance. Too frequently the four-fold table of results presents percentages only, without including the original raw frequencies. The following example serves to illustrate the importance of this omission. Suppose that, in an experimental population of 1,000 cases, a cutting-score is established which predicts failure for 60% of the men subsequently rejected, at a cost of only 10% of those who are acceptable in terms of the final criterion. If, however, only 50 of the 1,000 men were finally eliminated, while 950 were successful, the test would have cost 95 (10%) of the acceptable men, in the process of detecting 30 (60%) of those who should be eliminated. The problem then is to determine how many normals one can afford to eliminate, in order to detect a large proportion of the defectives. The answer to such a question can of course be determined only in the light of other factors, such as the seriousness of allowing defectives into the work, and the availability of men for the work in question. Where the unsatisfactory men may ruin expensive, complex, and difficult-to-replace equipment, or endanger the lives of other men, a large selection cost may be justified. In other situations where the manpower supply is tight, the selection cost must be reduced.

In some of the extensive work which has been undertaken by those working on selection procedures with the armed services, certain methods are being followed which may subsequently be judged to have been less than perfect. A few of the possible sources of error may be suggested. The tests being developed in many cases are still covering a more heterogeneous field than many feel desirable. Relatively pure tests, i.e. homogeneous tests, allow themselves to be subjected to a more rigorous interpretation. If we are to believe general reports, we find that the Army general classification test contains several quite different types of subject matter, verbal, mathematics, and spatial, and yet only a single score is obtained. The Army-Navy College Qualifying Test used as the first screening for over half a million men applying for the Army and Navy college programs contained three separate sections, verbal, mathematical, and science, and yet a single total score served as the basis for the screening. Other illustrations could be supplied. There are good practical and theoretical reasons for these complex tests being used, and for only a single score being reported; but a good case can also be made for the use of separate scores for each of the several components. It is interesting to note that revisions are now being developed for more pure types of measures in the Army basic battery.

---

[2] As here used, "selection cost" refers to the number of men falsely rejected by the selection procedure. The smaller the number of acceptable men rejected, the lower the selection cost.

In a few cases the selection programs use too many similar tests. In some cases, fifteen or sixteen test scores enter into the final selection. Possibly a reduced program of five or six relatively pure measures of meaningful complexes would do practically as good a job. Paper and pencil tests of the aptitude variety have their limitations, and the use of a large number of tests gives no assurance that the traits being measured are not few in number. Simpler programs of tests are indicated in some cases.

Another deficiency has been the lack of careful analysis of the results from the total testing program used. Such an analysis would reveal the weakness just mentioned. Factor analysis, for example, might be brought into play to show how many different factors are being measured by the tests employed. Factor analysis techniques, while not extensively used in the wartime selection-jobs known to the speaker, have been used effectively in several cases. Simple reliability analyses will show the undependability of certain of the measures with short time-limits still being used.

A serious error occasionally made is the establishment of selection techniques on populations not representative of the population for which the techniques are subsequently to be used. In many instances it can be demonstrated that the shifts in population are of great importance and the results of the tests cannot be interpreted without reference to the characteristics of the population being tested.

The acceptance of defective and unanalyzed criteria as the basis for the validation of tests and selection procedures constitutes a serious source of error. Ratings, service school grades, scores on performance trials—in short almost every available criterion measure—should be carefully checked and analyzed before it is accepted. How was the criterion measure obtained? Of what factors is it composed? Does it reflect the abilities and the skills which the tests were designed to measure or which are essential for success in the job? As testing programs become more carefully developed, as more homogeneous or pure types of measures are used, the final criterion also must be subjected to analysis and purification. In many cases it will be found inadequate and will impose a limitation on the interpretation of any validity coefficients based on it.

The poor conditions under which the tests must sometimes be administered also account for unsatisfactory results. At times the tests are administered to the men when they are in a frame of mind scarcely conducive to obtaining a normal sample of their behavior. For example, at one location scheduling complications led to the administration of aptitude tests to men immediately following inoculations. In another case, men were tested in the evening of their first day of very strenuous active work on a new location. Quite regularly, unselected men are being tested in groups as large as 500 or 1,000—some experts consider that better results could be obtained if groups were smaller. As more and more trained testing men take charge, improved procedures can be expected.

In spite of all these difficulties and all the weaknesses in the systems being used, results of demonstrable value are being produced. Relatively simple selection procedures are being shown to be of value in saving time and manpower, in putting the men in the jobs for which they have aptitude, and in eliminating the unstable and discontented from certain types of crucial work.

In looking toward the future, the growing intricacies of the machinery of war suggest that the country would be safer, if, directly in conjunction with the developments and techniques, more time were spent on research in methods of selection and training personnel. With new machines and improved techniques for the selection and training of men in their use, we shall be able to hold our own in any future situation. Psychologists and statisticians will do well to establish ever more clearly the gains to be obtained by simple but thorough selection procedures and the well-systematized use of test results in the armed forces.

Received October 2, 1944.

Adapting the Minnesota Rate of Manipulation Test to Factory Use

Guy M. Wilson and Staff

Personnel Testing Department, Raytheon Manufacturing Company,
Newton, Massachusetts

"Rate of movement is a unit skill and in and of itself cannot be improved. Only the techniques of performance can be improved." So states the author of the Minnesota Rate of Manipulation Test.[1]

If this is true, the measurement of an operator's rate of manipulation should reveal valuable information. It should provide a significant ranking of operators on the one item, speed of manipulation.

At one plant[2] the Minnesota Rate of Manipulation Test was included in a battery of tests in which it was sought to measure operators as to: (1) Intelligence—three tests, (2) Manipulative skill—three tests, (3) Special skill—two or three tests, according to the job. It appeared to hold its place as a helpful test under manipulative skill.

In time, however, some questions arose with reference to how best to use the test, and how to record the results. Time is always a factor in a production plant. Therefore the question—"Could we save two minutes, more or less, by using three trials instead of four?" The Manual for the Minnesota Rate of Manipulation Test calls for four trials and the final index used is the total time for the four trials. If three trials would serve as well, valuable time would be saved.

As the above question was studied, another question arose, viz., "Would the low score of four trials or three, serve as well as the sum of four trials or three, as an index?" If so, time would be saved in adding and a simple, more easily interpreted number could be used as the index. For example, the sum of four trials for an individual (see Table 1, which follows) might be 235 seconds. For the same individual the low of four is 54 seconds. The individual who sees this smaller figure, readily interprets it. It means, "One trial required 54 seconds."

The statistician knows that regardless of the convenience or reasonableness of a change in procedure, the change cannot be made unless statistics justify the change. In the case of these two questions the procedure for study was very simple.

---

[1] Zeigler, W. A. Manual for Minnesota Rate of Manipulation Test. Educational Test Bureau, Minneapolis.

[2] The Raytheon Manufacturing Company, Newton, Massachusetts.

Table 1 shows in column 1, the ordered arrangement of scores made by 63 subjects according to "sums of four." Column 2 shows the corresponding sums of three trials. Column 3 shows the corresponding lows of four trials, and column 4 shows the corresponding lows of three trials.

Table 1

Various Scores for Each of Sixty-three Factory Workers on the Minnesota Rate of Manipulation Test

(1) Sum of Four Trials	(2) Sum of Three Trials	(3) Low of Four Trials	(4) Low of Three Trials	(1) Sum of Four Trials	(2) Sum of Three Trials	(3) Low of Four Trials	(4) Low of Three Trials
177	124	40	40	229	174	55	56
184	137	40	40	229	174	55	56
192	141	44	44	230	176	54	55
196	148	48	48	230	175	55	57
199	149	49	49	231	172	56	56
200	151	47	47	232	178	54	55
203	150	49	49	235	176	54	54
206	152	48	48	235	182	53	58
209	158	49	49	235	173	54	54
209	155	50	50	236	178	58	59
210	158	51	51	239	182	57	60
211	159	52	53	241	183	58	59
212	158	52	52	243	181	59	59
212	160	50	50	243	183	60	61
216	165	51	53	244	184	59	59
216	162	53	53	245	182	59	59
216	163	53	53	246	186	59	59
216	162	52	52	247	182	60	60
216	165	51	54	248	185	61	61
218	165	53	53	251	189	60	60
218	162	52	52	253	189	62	62
218	163	52	52	255	192	60	60
219	166	53	54	255	191	60	60
220	164	54	54	255	184	59	59
221	167	54	55	255	194	61	62
223	171	52	55	256	194	61	61
224	170	54	55	256	193	63	63
225	169	55	55	258	197	62	65
225	170	55	56	261	196	64	64
226	168	55	55	263	199	64	66
226	171	55	55	280	219	61	67
228	169	56	56				

The present problem is so simple that the mere arrangement in order of the items, almost answers the questions raised. There are 15 or 16 misplacements in column 2, when compared with column 1, but the misplacements are small.

When column 3 is compared with column 1, the story is almost identical. There are 16 or 17 misplacements, all very small. In other words, the low of four trials would give almost the same ranks as the sum of four trials. And the same applies to column 4, the low of three trials.

If three trials are as good as four, or approximately so, and if the low of the trials is as good as, or better than the sum, then we can move on to the *low* of *three* trials, as the index to use. It will save time, and it will be easily understood.

Comparing columns 3 and columns 4, the low of three trials and the low of four trials, it appears that for 43 of the subjects there is a zero difference. For instance, the first case, the low of four is 40 and the low of three is 40,—in other words, no difference. In 12 of the pairs the difference is one. For instance, the first difference is the twelfth case, the low of four is 52, the low of three is 53. There are 12 such pairs as indicated above where the difference is one. Thus we have 43 plus 12 cases in which the difference is zero or one. The other differences are as follows: 2 cases with a difference of 2; 4 cases with a difference of 3; 1 case with a difference of 5; and 1 case with a difference of 6.

The author of the Minnesota Rate of Manipulation Test does not present the data supporting the reasons for the choice of the sum of four trials as the proper index. The sum, of course, is equivalent on a ranking basis to the average. An average would give a lower figure, and, therefore, a more easily comprehensible index. On theoretical grounds, and in the absence of supporting data, it may be easily argued that a low of four trials is better than an average of four trials. In the field day event, such as the pole vault or the high jump, the best score made is taken, not the average.

It is evident, from a casual study, that the correlation between any two columns in Table 1 is very high. The correlation between columns 3 and 4, for instance, using the product moment formula, is + .97. The only negative product in the products column is − 3; there are four zeros; the other 58 products are positive.

Correlating[3] the other columns of Table 1, gives the following values for r—columns 1 and 2, + .986; columns 1 and 3, + .952; columns 1 and 4, + .968; columns 2 and 3, + .939; columns 2 and 4, + .968.

It was finally concluded in this particular factory to substitute the low of three trials for the sum of four trials as the index of performance for the Minnesota Rate of Manipulation Test. It is more convenient; it is more easily understood by some one to whom an explanation of the score is being made. It, probably, is an equally good index, although the proof of this last statement would require checking by more cases than used in this study and correlating with outside criteria. But this study is sufficient to raise the question, and, probably, to justify the change to a more convenient and understandable index.[4] In a busy factory, time and ease of understanding are important factors.

---

[3] Correlations figured by Rachel Lounsbury.

The above discussion may lead the reader to suspect the use of local data for the establishment of local norms. This is correct. First interpretations were based upon national or published norms. As soon as sufficient cases were at hand to fill in what appeared to be a typical distribution, local norms were tentatively established. If confirmed by later distributions, they were then used with reasonable confidence. Constant checking of one's data is necessary in any case and such checking sometimes reveals desirable local adaptations.

Received October 7, 1944.

---

[4] See also Jacob Tuckman: A comparison of norms for the Minnesota Rate of Manipulation Test. *J. Appl. Psych.*, 28: 121–128, Apr. 1944.

The Horn Art Aptitude Inventory

Charles A. Horn and Leo F. Smith

Rochester Institute of Technology, Rochester, New York

It is the purpose of this paper to describe the Horn Art Aptitude Inventory. This aptitude test has been developed by the faculty of the School of Applied Art of the Rochester Institute of Technology[1] during the past eight years and has been used with freshmen entering the Art School at this institution and with groups of high school art students competing for Art School Scholarships.

Construction

After having studied and experimented with various art tests over a period of years, the Art School faculty were of the opinion that certain qualities essential to success in the art field were not being satisfactorily measured. The problem of obtaining clues to these qualities in students was the objective in designing this test.

The test is divided into two distinct sections: (1) Drawings of Lines and Shapes, subdivided into two parts: (*A*) Scribble Exercise, and (*B*) Doodle Exercise; and (2) Imagery.[2]

In section 1 (Part 1*A*) the Scribble Exercise is designed to give the student confidence that he can draw a reasonably simple shape or picture. In this part he is asked to draw twenty different items such as a book, a fork, etc. and is given a limited time, varying from two to six seconds, in which to make each drawing. The total time required for administration of this section is approximately five minutes.

The Doodle Exercise (Part 1*B*) is designed to obtain examples of the student's quality of lines, ability to follow directions, originality and compositional sense. In this part he is asked to draw various lines and shapes such as rectangles, triangles, circles, etc. The total time required for the administration of this section is approximately five minutes.

The Imagery Section (Part 2) is designed to obtain an indication of the scope of the student's interests, and the fertility of his imagination with respect to the number of ideas and the ability he exhibits in presenting these ideas. In this section there are twelve rectangles 2¾ inches by 3½ inches in which certain key lines are presented and the student is requested to use these lines as "spring boards" and construct sketches which are suggested to him by the key lines. Figure 1 shows two pictures which have been constructed using the key lines given. The total time required for the administration of this section is approximately forty minutes.

---

[1] Formerly Rochester Athenaeum and Mechanics Institute.

[2] The test and manual of directions is distributed by Educational Research Office, Rochester Institute of Technology.

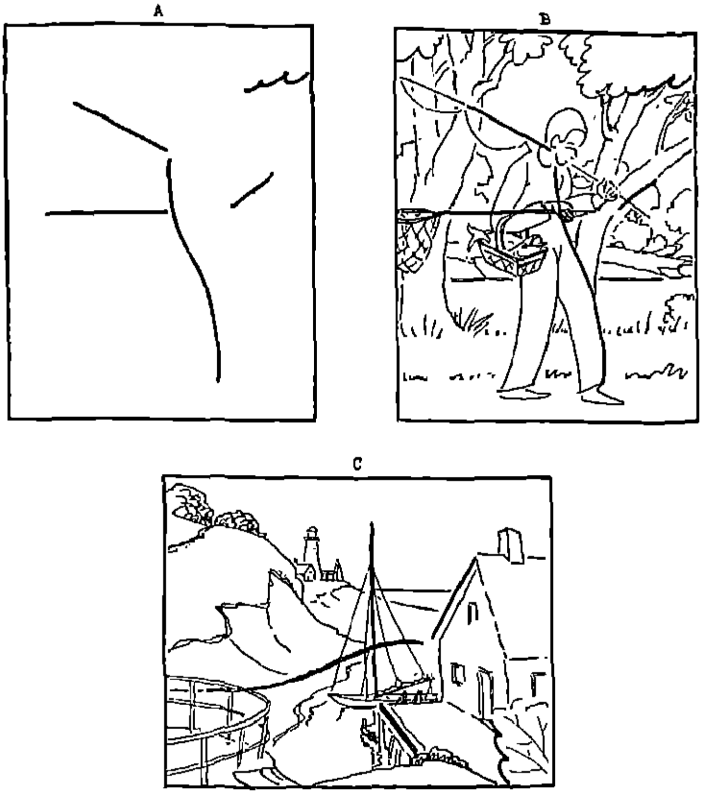


Fig. 1. Sample pictures B and C drawn on basis of key lines shown in A.

Interpretation and Scoring

In the manual of directions now being prepared for use with this test examples of excellent, average, and poor papers are being included. This will enable a person not trained in art to have a basis upon which to make judgments. There are several standards, however, which the members of the Art School faculty have identified as important. These are:

1. *Order:* Are the items so placed on the sheet that they fill it pleasingly and indicate that the student has a sense of order, or has the student

cramped the items so that the entire page was not used? In other words has the work been well planned for the size sheet which is being used?

2. *Clarity of Thought and Presentation:* Are the sketches made with a clean line so that there are no erasures or fumbling? Are the items recognizable, that is, if the student meant to draw a tree does the drawing reasonably resemble a tree? What is the quality of the line used? Is it continuous? Is it broken, cramped, or bumpy or is it graceful and smooth?

3. *Color:* Is there a consistently even tone to the drawings or does the sheet appear spotty, i.e. is there evidence of uneven pressures: Are the items smudged, fuzzy or erratic in their line quality?

Interpreting the Scribble Exercise 1-A: The directions provided with the scoring manual are as follows:

Open up the folder to page 2 and fold the top sheets back. Lay the tests side by side on a long table so that you have an over-all view of the work of all the papers.

Scan up and down the papers quickly keeping in mind the standards of order, clarity and color. In this manner those which are "excellent" and those which are "poor" may be readily identified and it will be noted that this generally leaves a group which may be considered "average." If it is desired to obtain a more precise judgment of rank than "excellent," "average," and "poor," by use of the criteria which have already been suggested, divide the "average" group into "good," "average," and "fair." This then gives five categories: "excellent," "good," "average," "fair," and "poor."

It is considerably more difficult to break up the "average" group into three sub-groups than it is to judge the extremes. It has been found that a lay person, without any training in art work, can judge the extremes with as great accuracy as can competent art teachers, but a lay person experiences somewhat more difficulty when efforts are made to divide the "average" group into the three sub-groups.

Interpreting the Doodle Exercise 1-B: One important student trait often identified here is originality. For example, has the student done the usual thing and divided the square and rectangle in a symmetrical manner, which would indicate conformity or triteness, or has he spotted the smaller square off center. Similarly, has he broken up the rectangle exactly in the center of the sides or has he done the unusual, i.e., broken it up in an asymmetrical manner?

Interpreting the Imagery Section—2: The criteria which have already been mentioned should be kept in mind but in addition the following should be noted:

1. The fertility of imagination as indicated by the number of ideas presented.
2. The scope of interests. Are these limited to one particular type such as landscapes, people, sea scenes, or does the student have a wide range of interests?
3. The clarity of mental image.
4. Color—Does the student utilize an outline only or does he shade much of his work?
5. Design—Does the student consistently use abstract forms as contrasted with the literal or naturalistic?

The junior author of this paper (L. F. Smith), who is a member of the Educational Research Office and has had no art training, has scored two groups of these inventories employing the technique of spreading the tests side by side on a long table and identifying those which are "excellent," "average" and "poor" on a subjective basis. Then, using the criteria which have already been suggested the "average" group has been divided into "good," "average," and "fair."

Table 1

Reliability of Scoring Horn Art Aptitude Inventory

[Coefficients of correlation between ratings given by two Art School Faculty members (A_1 and A_2) and member of Educational Research Office (E.R.O.)]

	Group I *				Group II **	
	A_1	A_2	E.R.O.		A_1	E.R.O.
A_1	—	.85	.86	A_1	—	.79
A_2		—	.83	E.R.O.		—
E.R.O.			—			

* Group I consisted of 21 Art School students who took this test in the Fall of 1939.

** Group II consisted of the Scholarship Class of 20 high school seniors who took this test during the Spring of 1944.

Table 1 illustrates the reliability of ratings for these two groups. In Group I two Art School faculty members and the junior author independently scored 21 test papers of regularly enrolled art school freshmen. In Group II one of the Art School faculty and the junior author independently scored 20 test papers of high school seniors competing for fellowships in the School of Applied Art.

Validity

Two studies of validity have already been made and others are in progress. In the first study all of the students who graduated from the

Art School in 1941, '42 and '43 were rated by four Art School faculty members on their success in the three year course. The Horn test scores of these students ($N = 52$) were then correlated with the average of the faculty ratings. The Pearson product-moment correlation between the Horn test scores of these students and the average faculty rating was + .53.

In the second study the 36 high school seniors enrolled in the Fellowship Competition Classes of 1943 and 1944 were rated on their success in this class by four Art School faculty members. The Horn test had been given to all of these students at the beginning of the class and the product-moment correlation between their scores and the average faculty rating of success was + .66.

That the Horn Art Inventory measures something other than intelligence is indicated by the product-moment correlation of only + .15 between the Inventory test scores for the classes of 1941, '42, and '43 ($N = 52$) and their American Council on Education Psychological Examination scores. That this inventory is of more value in predicting success in the three-year art course than is the A.C.E. intelligence test is indicated by the product-moment correlation of + .28 between the intelligence test scores and the average faculty rating of success for these same three classes.

Summary

1. The Horn Art Aptitude Inventory has been in the process of development for a period of more than eight years at the Rochester Institute of Technology.

2. The unique features of this Inventory are: (*A*) The student is required to make reasonably simple drawings which illustrate the quality of line he employs, his appreciation of proportion, and his compositional sense, and (*B*) the student is given exercises which provide an indication of the scope of his interests, the fertility of his imagination, and the ability to depict pictorially ideas which occur to him.

3. The scoring is still somewhat subjective but correlations between the ratings given test papers by Art School faculty members and a lay person vary from .79 to .86 for two different groups of students. It appears that a lay person with no training in art can score these papers as adequately as members of the Art School faculty.

4. The product-moment correlation between the Horn Inventory test papers and faculty rating of success of the 52 graduates of the classes of 1941, '42 and '43 in a three-year full-time program was + .53. The correlation between test scores and success in two much shorter Scholarship Classes was + .66 ($N = 36$).

5. The correlation between scores on the Horn Inventory and A.C.E. intelligence test scores is low (+ .15). This Inventory is of more value in predicting success in the three-year Art School course than is the A.C.E. intelligence test as the correlation between the latter and course success was + .28.

Additional studies are being carried on in the effort to make the scoring of test papers more objective and to determine the effectiveness of this as a predictive instrument for different age groups. It is believed that the results of these studies will improve this instrument which has already been of value in one institution.

Received September 18, 1944.

A New Method for the Administration of Individual Intelligence Tests

Raymond Corsini

Auburn Prison, Auburn, New York

Most test manuals and books on tests and measurements agree rather well on the general methods of conducting an individual intelligence test (1) (2) (3) (4) (5) (6) (7) (8) (9). Terman and Merrill (7) summarize in three directives: "(1) Standard procedures must be followed, (2) the child's best efforts must be enlisted by the establishment and maintenance of adequate rapport, (3) responses must be correctly scored."

However, there is a lack of directions in any manual for the actual administration of an individual test in terms of three variables: (1) where to place subject in relation to examiner, (2) where to keep test materials during the course of the examination when not in actual use, and (3) how much of the behind the scene actions of the examiner are to be permitted to be seen by the subject.

The purpose of this article is to give a description and evaluation of various ways in which these three variables are met by examiners, plus the description of a new method for administering individual tests which appears to be superior to any in present use.

Placement

Generally the subject (1) sits at the right hand side of the examiner's desk, (2) or faces the examiner behind a table. The second method is more comfortable for the subject if he has any writing to do, or if he has to handle any material.

Materials

Some examiners stow all materials in a desk drawer. Some keep materials in small boxes, within a larger box which in turn is put on the desk. Some scatter materials loosely over the desk or table. Some examiners keep all material out of sight except when in use. Others permit material to accumulate on the desk or on the table.

The best method appears to be that which is most convenient for the examiner and which permits no wasteful searching around for an article. Generally, it seems best to handle any items so that they will not distract the subject.

Scoring

There are four popular ways to score the test blank. Two are done in the subject's sight, two are done out of sight.

Method "A" scores openly on the desk or table.

Method "B" scores openly but makes a check mark to indicate correct, and a check mark with a loop to indicate wrong.

These methods have the good point that they help keep up rapport in that nothing is hidden. Method "A," however, is poor because subject tends to change or add to his answer on getting a minus, or may even demand why he is being scored incorrectly.

Method "B" may fool some very dull subjects, but generally subjects know when they are wrong, and realize that some hidden procedure is in operation. This method tends to cause unrest on the part of the subject.

Methods "C" and "D" involve scoring the protocol out of sight. In method "C," the best of these four methods, a visual barrier made either from a folder or made of more permanent material is interposed on desk or table between subject and psychologist. Behind this barrier, the psychologist prepares materials, and scores the protocol. The advantage is in not letting the subject see what he is being marked, but its disadvantage lies in its abruptness and the "insult" to the subject.

Method "D" involves folding the test booklet into quarters, keeping it flat on the bottom of a desk drawer, together with stop watch and manual, and attempting to mark the protocol in an unobstrusive manner with a stub of a pencil. This method, to the author, seems the worst of the four, since it soon becomes obvious to the subject that the examiner is reading from a book in the desk, and is slyly making notes meanwhile.

The New Method

For some time the author of this article has followed a novel procedure in administering individual tests that appears to possess superior advantages to any of the combinations of the three general variables so far described.

Following the interview, the subject is asked or is told to take an individual test. A table, 18″ × 30″, is placed parallel to the pull-out leaf on the right hand side of the examiner's desk. The subject sits behind the table, facing the examiner. From the subject's point of view a box is then placed on the upper right hand corner of the small table.

This arrangement allows the subject to sit behind the table, with his feet under it. He has plenty of space to write. The examiner makes his notations on the pull-out leaf of his desk.

As soon as both are settled, the examiner removes the test manual and a test protocol (blank for scoring responses) from the box. This immediately indicates the function of the box to the subject. The manual is placed on the desk in view of the subject but too far away for him to be able to read from it. The protocol is placed on the pull-out leaf in front of the box, therefore out of sight of the subject. The box acts as a visual barrier, but in so natural a manner that it cannot disturb a subject since he accepts the box as an integral part of the examination procedure.

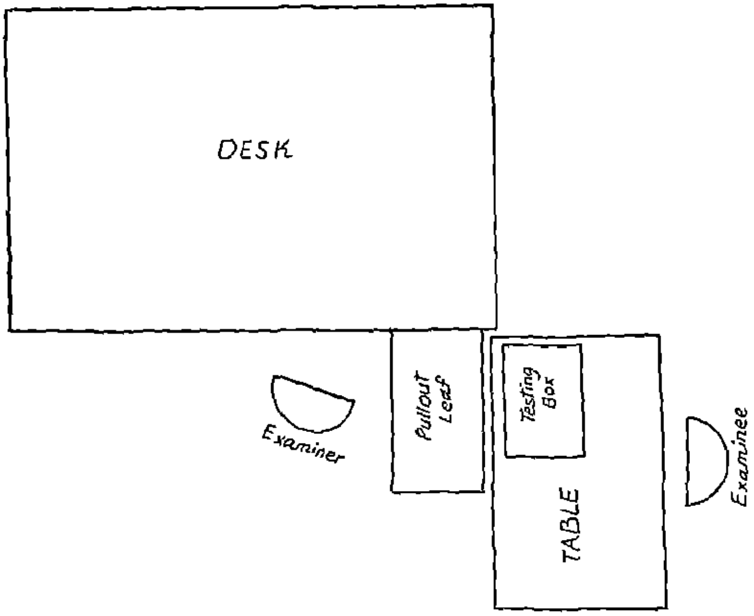


Fig. 1. Schemata for individual test administration.

Whenever any test materials are needed, such as blocks, cards, tissue paper, scissors, form boards, etc., they are taken from the box, quickly and simply, since they are found in various compartments of the three shelves or sections, and immediately after use are returned to their proper places. Everything is instantly available, nothing can accumulate, and since there is no searching for materials the examination proceeds swiftly and efficiently.

The author has had two such boxes constructed, one for the Wechsler-Bellevue and one for the Stanford-Binet. Of course, the arrangements of the partitions vary to accommodate different materials. The boxes are uniform in size, approximately 5″ × 10″ × 12″. Each box consists of three elements or shelves, stacked on top of each other, hinged at the back, and locked to each other at the front by a simple catch. The top element has a hinged cover.*

At the conclusion of the test, the manual is replaced, and the box is put away, ready for the next administration.

Summary

1. There is no uniformity of individual test administration with respect to these variables: *a.* Placement of subject and examiner; *b.* Maintenance of material during the course of the examination; and *c.* Scoring the test blank.

2. A new method of administering individual tests is described which possesses the following advantages: *a.* Is standard; *b.* Is fast, efficient, and simple; and *c.* Reduces subject-examiner friction.

Received September 22, 1944.

References

1. Burt, C. *Mental and scholastic tests.* London: P. S. King and Sons, 1922.
2. Gregory, C. A. *Fundamentals of educational measurement with the elements of statistical method.* New York: D. Appleton and Co., 1922.
3. Kohs, S. C. *Intelligence measurement.* New York: MacMillan, 1927.
4. Kuhlman, F. *A handbook of mental tests.* Baltimore: Warwick and York, 1922.
5. Mateer, F. *The unstable child.* New York: D. Appleton and Co., 1924.
6. Porteus, S. D. *Guide to Porteus maze test.* Training School, Vineland, New Jersey, 1920.
7. Terman, L. M., and Merrill, M. A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
8. Wechsler, D. *The measurement of adult intelligence.* Baltimore: Williams and Wilkins Co., 1939.
9. Wells, F. L. *Mental tests in clinical practise.* Yonkers: World Book Co., 1927.

---

*These boxes are obtainable from C. H. Stoelting Co., Chicago.

The Relationship Between Scholastic Achievement and Personality Adjustment of Men College Students

George R. Griffiths

Division of Personnel Management, U. S. Maritime Commission

Personality is often considered an Alladin's lamp to achievement. If one has "personality," success is inevitable. If one has no "personality," he may as well resign himself to his fate. Fortunately, such an attitude is fast being displaced by a scientific appreciation of the true nature of that phantom "personality."

Broadly speaking, personality consists both of a person's reactions and responses and of the influence that person wields over others. Space limitations preclude a lengthy discourse on the nature of personality, but let us take a moment to view some of its main factors or traits. We can do no better than to refer to Allport's table of personality traits (1). Listed concisely, they are: 1. Intelligence; 2. Motility; 3. Temperament; 4. Self-expression; and 5. Sociality.

Some authors add a sixth, *physique*. Gaskill (3), over a ten-year period, made a survey in his beginning psychology classes of the traits most desired in a mate. Seventy per cent ranked health first; forty-five per cent ranked intelligence second. His list of the elements of personality is as follows: 1. Intelligence; 2. Social adjustment; 3. General characteristics of overt behavior; and 4. Physical characteristics.

A scientific analysis of personality, its traits, and its relationships, should proceed by the examination of specific factors and specific relationships. The problem here undertaken is to determine whether or not there is a significant relationship between personality adjustment and academic achievement.

Previous Investigations

Those who blandly state that intelligence does not correlate with personality are ignoring the fact that intelligence is an integral part of personality. What these persons do mean, however, is that intelligence does not correlate highly with various other personality traits. High intelligence, according to Strang (9), is ordinarily associated with a pleasing personality, since intelligence involves insight, the ability to see relationships, and the capacity to learn.

A brief summary of various other pertinent studies may serve to reveal the nature and findings of previous investigations. A substantial positive relationship between intelligence as measured by standardized tests and personality as judged by interviews and ratings was found in University of Iowa studies (2). However, when scores on personality tests and questionnaires are used rather than the results of observation, the coefficients of correlation range around zero (between + .20 and − .20).

Terman (10) found that superior children (most of them with IQ's above 140) are more emotionally stable and more socially adequate than unselected children. One study (4) showed a tendency toward small negative correlations between scores on neurotic inventories and scholarship. On the other hand Thurstone found, in applying his Personality Schedule to college students, that no relationship existed between intelligence and "neurotic tendency," but that the less well-adjusted students tended toward slightly higher academic grades (6). There is a tendency for higher learning performance to be associated with submissiveness as measured by the A-S scale (5).

Stagner's findings (7) were that unstable and maladjusted students do less well in proportion to their intelligence than do stable persons; that introverts earn proportionately higher marks; and that unfavorable scores in emotionality and self-sufficiency are associated with lower achievement than would have been predicted from intelligence alone. Strang (9) reports a lack of relationship between scholarship and various measures of introversion-extroversion. Finally, reference should be made to a study by Stedman (8). She found that the average grades of pupils with certain health defects were only 75% those of healthy pupils. However, of her 450 cases, the healthy group numbered only 39.

Occasionally studies appear to show definite relationships between personality and scholarship. Others seem to be contradictory. The result is that there has as yet been no clear definition of the connection between scholarship and personality.

Study at Ohio University

This study was undertaken with Freshman men at Ohio University to discover whether there is a relationship between scholastic achievement and, personality adjustment, employing the statistical technique, the probable error of the difference between means. The measure of scholastic achievement used was the first semester point-hour-ratio; i.e., the total number of semester hours of courses carried divided by the number of points earned, where $A = 3$, $B = 2$, $C = 1$, and $D = 0$. The Bell Adjustment Inventory was used as a measure of personality adjustment. This Inventory measures four types of personality adjustment: health, home, social and emotional, yielding scores for each area and a total, composite score. (Strang is firm in her assertion of the unreliability of such personality questionnaires (9). Her argument is based mainly on the fact that they fail definitely to differentiate psychiatric patients. She may be right; that is not the problem here.)

In this study of personality and scholastic achievement several different viewpoints were utilized. The first approach was to compare with various other groups those men placed on scholastic probation at the end of their first semester for earning a point-hour-ratio of less than 0.5. The other groups are: first, a group matched person for person with the Probation group in college ability scores; i.e., scores on the Ohio State University Psychological Examination (Matched group); second, a group selected at random (Average group); and third, a group matched individually with college ability scores as high as those of the Probation group were low (Excellent group).

Table 1

Comparison of Probation Group with Other Groups in College Ability, Grades, and Personality

Group	*N*	College Ability Percentile	Point-Hour-Ratio	Total Personality Score
Probation	40	21.3	0.260	30.2
Matched	40	21.3	1.023	36.4
Average	40	40.8	1.410	37.8
Excellent	40	78.7	1.703	34.8

Statistical comparisons of these groups in college ability, grades, and personality (as measured by the Bell Adjustment Inventory) are presented in Table 1. All figures are arithmetic means. A low score is the favorable score on the Bell Adjustment Inventory; so the lower the mean, the better. Probable errors of the difference between means were computed as a measure of the significance of those differences. It is generally accepted that to be statistically significant the difference should be at

Table 2

Probable Errors of Mean Differences in Personality of Probation and Other Groups

Groups	Difference of Means	P.E. (Diff.)
Probation and Average	1.4	2.367
Probation and Matched	2.8	2.153
Probation and Excellent	4.4	1.904

least four times its probable error. The difference in mean personality scores and the probable errors are shown in Table 2. In no instance is the difference significant since the probable errors are too great. In the comparison of the Probation men with Excellent men the greatest difference is found; but it is only 2.2 times its probable error. Apparently, men experiencing scholastic difficulty exhibit no significant personality differences from persons of superior college ability.

Another comparison made was of the Bell Adjustment Inventory scores of Probation men and men matched as nearly as possible with point-hour-ratios as high (Opposite group) as those of the Probation group were low. Table 3 includes scores of these two groups in college ability, grades, and personality. The difference of 1.8 between the mean personality scores is clearly not significant, as the probable error of this difference is 2.099. In other words, men on scholastic probation are approximately equal in personality adjustment to men with superior scholastic records.

Table 3

Comparison of Probation and Opposite Groups in College Ability, Grades, and Personality

Group	*N*	College Ability Percentile	Point-Hour-Ratio	Total Personality Score
Probation	40	21.3	0.266	39.2
Opposite	40	85.3	2.610	37.4

Table 4

Comparison of Groups Divided on the Basis of Personality Scores

Group	*N*	Total Personality Score	College Ability Percentile	Point-Hour-Ratio
Very unsat.	37	66.9	42.3	1.018
Unsatis.	122	48.8	39.5	1.004
Average	77	38.5	48.0	1.198
Good	112	17.5	49.0	1.169
Excellent	17	7.1	50.7	1.263

A third approach was made by comparing groups of men divided on the basis of their total scores on the Bell Adustment Inventory. Five, groups are thus differentiated: very unsatisfactory, unsatisfactory, average, good, and excellent. In Table 4 are presented averages in personality, college ability, and grades. The difference in point-hour-ratio between the very unsatisfactory group and the excellent group is 0.245, in favor of the excellent group. The probable error of this difference is 0.1187. Since this figure is less than one-half the difference, the difference between the means of these two groups is not statistically significant. Apparently, men students scoring very unsatisfactory on the Bell Adjustment Inventory do not reveal scholastic trends significantly different from those scoring excellent in personality.

The fourth attack on the problem was a comparison of two groups selected on the basis of health scores in the Bell Adjustment Inventory. This viewpoint was suggested by the study of Stedman cited above. Men scoring very unsatisfactory in health were separated from those scoring excellent. Table 5 contains their mean scores in personality, college ability, and grades. As an interesting sidelight, the grades of those with very unsatisfactory health were 78.6% of those in excellent health. Although based on smaller samples, this appears to be in line with Stedman's finding of 75%. However, the difference between the mean point-hour-ratios is 0.308, which is only 2.3 times its probable error, 0.133, and, therefore, not significant. We can conclude, then, that men scoring very unsatisfactory in health are not particularly inferior in scholastic achievement to, although a difference appears to exist in favor of, those in excellent health.

Table 5

Mean Scores of Groups Divided on the Basis of Health

Health	*N*	Total Personality Score	College Ability Percentile	Point-Hour-Ratio
Very unsat.	25	53.1	50.3	1.133
Excellent	22	17.7	55.6	1.441

A fifth comparison made was of the grades of men scoring very unsatisfactory and men scoring excellent in emotional adjustment on the Bell Adjustment Inventory. Table 6 contains mean personality scores,

Table 6

Mean Scores of Groups Divided on the Basis of Emotional Adjustment

Emotional Adjustment	*N*	Total Personality Score	College Ability Percentile	Point-Hour-Ratio
Very unsat.	40	58.0	42.58	1.144
Excellent	40	14.9	56.65	1.118

college ability percentiles, and point-hour-ratios of these two groups. The difference in point-hour-ratios is 0.026 in favor of the very unsatisfactory group. This is the only difference in the study which appeared contrary to expectation. It corresponds, however, with the tendency Thurstone found in studies mentioned above. As its probable error amounts to 0.1001, no importance can be placed on the difference. The difference is also the smallest revealed in the study as it is but one-fourth its probable error.

A final analysis utilized comparison of the highest and lowest deciles in college ability to see whether differences in personality might exist. This is more nearly a comparison of personality with intelligence. Mean scores for these two groups are contained in Table 7. In this case the difference in personality scores is 5.0 in favor of those in the highest decile. This difference is 2.1 times its probable error and, hence, not great enough to be accepted as of statistical significance. The only conclusion that can be drawn is that men in the highest and lowest deciles of college ability do not show a marked difference in personality.

Table 7

Mean Scores of Groups Divided on the Basis of College Ability Scores

Group	*N*	Point-Hour-Ratio	Total Personality Score
Lowest 10th	39	0.533	42.7
Highest 10th	38	2.092	37.7

Results and Conclusions

The question of whether there are valid relationships existing between scholastic achievement (point-hour-ratio) and personality (Bell Adjustment Inventory) has been examined here from several different points-of-view. The results are these, briefly:

1. Men clearly in scholastic difficulty, having been placed on academic probation, are not very much inferior in personality adjustment scores to men of superior college ability (Tables 1 and 2).

2. Men students with brilliant scholastic records are no better adjusted in personality than men of lowest academic achievement (Table 3).

3. An analysis of men with very unsatisfactory personality scores shows no significant difference in their grades from those with excellent personality adjustment scores (Table 4).

4. Scrutinizing a comparison of men with very unsatisfactory health scores with men of excellent health scores reveals a small but not significant difference in favor of the excellent group (Table 5).

5. Men with very unsatisfactory emotional adjustment scores tend toward higher grades than men of excellent emotional adjustment scores, but the difference is not significant (Table 6).

6. There is no very great difference in personality scores evident between men in the lowest decile of college ability (The Ohio State University Psychological Examination) and men in the highest decile (Table 7).

In every case but one there is such a difference as suggests some degree of positive correlation between scholastic achievement and personality. But as a difference, to be accepted as statistically significant, must be at least four times its probable error, the differences found are not large enough to be valid. In these analyses the differences ranged from 0.25 to 2.3 times their probable errors. Nevertheless, it seems reasonable to conclude that the consistency of these differences, even though they are small, is in itself important. It may mean that our psychometric techniques, especially personality measures, are in need of refinement. Then, too, it may mean that actual differences do not exist, however logical it is to expect them. In any event, no conclusions can be safely drawn until further research is conducted with more positive results.

Suggestions for Further Study

Further study along two lines might be productive of useful results:

1. The difference in mean personality scores between the Probation and the Matched groups in favor of the latter (Table 1) hints that personality factors may be present to influence the difference in grades of persons of equal college aptitude. The causes of diverging academic records of persons of approximately equal mental ability should be investigated to determine whether personality factors are present.

2. Since college students are highly selected, nearly all being above normal in intelligence, studies should be made where groups definitely below average can be compared with those high in intelligence.

Received September 30, 1944.

References

1. Allport, Floyd H. *Social psychology.* Boston: Houghton Mifflin, 1924.
2. Francis, Kenneth V., and Fillmore, Eva A. The influence of environment on personality of children. *Univ. Ia. Stud. Child Welf.*, 1934, 9, 71.
3. Gaskill, Harold V. *Personality.* New York: Prentice-Hall, 1936.

4. Hertzberg, Oscar E. Emotional stability as a factor in teachers college admissions and training. *Educ. Adm. Supervis.*, 1933, 19, 141–148.
5. McGeoch, John A., and Whitely, Paul L. Correlations between certain measurements of personality traits and of memorizing. *J. educ. Psychol.*, 1933, 24, 16–20.
6. Shaffer, Laurance F. *The psychology of adjustment.* Boston: Houghton Mifflin, 1936.
7. Stagner, Ross. The relation of personality to academic aptitude and achievement. *J. educ. Res.*, 1933, 26, 648–660.
8. Stedman, Melissa B. The influence of health upon intelligence and school grades of high school pupils. *J. appl. Psychol.*, 1934, 18, 799–809.
9. Strang, Ruth M. *Behavior and background of students in college and secondary school.* New York: Harper, 1937.
10. Terman, Lewis M., and Others. Mental and physical traits of a thousand gifted children. *Genetic studies of genius,* Vol. I (2nd Ed.), Stanford Univ., Calif.: Stanford Univ. Press, 1926.

Negro-White Attitudes Towards the Administration of Justice as Affecting Negroes

F. C. Sumner and Dorothy L. Shaed

Howard University

It was proposed in this study to measure the degree of unanimity in attitudes of Negroes and whites both at the college level and at the adult level with respect to the administration of justice as affecting Negroes.

Method

A questionnaire was devised consisting of 56 statements taken from the spontaneous conversations of Negroes. The respondents were instructed to read each statement and to indicate their reaction with a circle in one of the following six ways:

If you feel that the statement is absolutely true, draw a circle around the symbol T_4.

If you feel that the statement is more true than false, draw a circle around the symbol T_3F_1.

If you feel that the statement is about equally true and false, draw a circle around the symbol T_2F_2.

If you feel that the statement is more false than true, draw a circle around the symbol T_1F_3.

If you feel that the statement is absolutely false, draw a circle around F_4.

In case you do not understand a statement, draw a circle around the question mark.

Personal information was requested such as sex, age, race and whether or not one had had any court experience (By court experience was meant any experience from being merely a "spectator" to being a judge). It actually turned out that the attitudes of those with court experience differed very slightly from the attitudes of those without court experience.

Of the 1,099 persons replying to the questionnaire there were 246 white college students of whom 176 were male and 70 female; 660 Negro college students of whom 261 were male and 399 female; 193 adults of whom 42 were white and 151 Negro.

The 906 college students replied from the following colleges:

University of Illinois	66	(M 44, F 22)
University of North Carolina	90	(M 81, F 9)
University of South Carolina	54	(M 31, F 23)
University of Florida	36	(M 20, F 16)
West Virginia State College	89*	(M 35, F 54)
Howard University	48*	(M 26, F 22)
University of Illinois	13*	(M 0, F 13)
Tennessee State College	97*	(M 26, F 71)
Florida A. and M. College	94*	(M 49, F 45)
Alcorn A. and M. College (Miss.)	77*	(M 53, F 24)
State College, Orangeburg, S. C.	105*	(M 27, F 78)
Virginia State College	137*	(M 45, F 92)

Negro college students starred.

The 193 adults who replied lived in the District of Columbia.

In reducing the great mass of raw data to manageable terms the following formula was devised and designated the True-False Index (TF Index) of a group in respect to a particular statement:

$$\frac{(T_4 + T_3F_1 + \tfrac{1}{2}T_2F_2) - (\tfrac{1}{2}T_2F_2 + T_1F_3 + F_4)}{N\ (=\text{ total number of replies to the specific statement})}.$$

For example, the 174 white college males replying to Statement No. 27 (Judges are entirely free of racial prejudice) distributed as follows:

T_4	T_3F_1	T_2F_2	T_1F_3	F_4
9	35	24	55	51

and the TF Index is

$$\frac{\left(9 + 35 + \frac{24}{2}\right)}{174} - \frac{\left(\frac{24}{2} + 55 + 51\right)}{174} \text{ or } 32\% - 68\% = -36\%.$$

This obtained TF Index means that 36 per cent voted against the proposition over and above the remaining 64 per cent who were tied between accepting and rejecting it.

TF Indices vary between + 100 (unanimous belief of the group in the truth of a statement) and − 100 (unanimous disbelief of the group in the truth of a statement). When TF Indices are + 100 to + 34 inclusive, they indicate a definitely positive reaction on the part of the group in as much as two-thirds or more of the group accept the proposition; when TF Indices are + 33 to − 33 inclusive, they indicate a definitely mixed reaction on the part of the group in as much as two-thirds or more of the group are tied between accepting and rejecting the proposition; when

TF Indices are − 34 to − 100 inclusive, they indicate a definitely negative reaction on the part of the group in as much as two-thirds or more of the group reject the proposition.

Results

Percentages of the 56 statements to which positive, mixed and negative reactions are made by each of the several groups are given in Table 1.

Table 1

Percentages of the 56 Statements to which Positive, Mixed and Negative Reactions are Made by Each of the Several Groups

Group	Positive	Mixed	Negative
All White Adults	7%	71%	21%
All White College Students	34	41	25
White College Males	32	43	25
White College Females	30	30	30
All Negro Adults	45	27	29
All Negro College Students	46	34	20
Negro College Males	41	34	25
Negro College Females	48	32	20

From Table 1 it appears that in the white adult group the percentage of mixed reactions is higher than either that of positive or that of negative reactions and even higher than the combined percentages of positive and negative reactions. In the white college groups the percentage of mixed reactions is higher than either that of positive or that of negative reactions but not higher than the combined percentages of positive and negative reactions. On the other hand, it appears that in all Negro groups percentages of positive reactions are higher than either that of mixed or that of negative reactions while combined percentages of positive and negative reactions are in every case higher than the percentages of mixed reactions.

The very strong tendency of the white adult group towards mixed reactions (two-thirds or more of the group being tied between accepting and rejecting the statements) may be thought due at least in part to the fact that the issuing of the questionnaires to this group was done in person by a Negro which may have in a selective or moderating way influenced the reactions. On the other hand, the mixed reactions of this adult white group appear to be but a fuller manifestation of a tendency to reservation, i.e., to mixed reaction already perceptible in every white group of college students despite white administration of the questionnaires. Factors more likely influencing white groups to mixed reactions may be gleaned to some extent from scattered comments written in the

Table 2

TF Indices of the Several White and Negro Groups for Each of the 56 Statements

	Adults	College Students	Male College Students	Female College Students
White	42	246	176	70
Negro	151	660	261	399
Total	193	906	437	469
1. The practice of Negro lawyers should be confined to routine office work.				
White	−55	−51	−53	−47
Negro	−97	−64	−69	−70
2. In the eyes of the court one white witness is better than any number of Negro witnesses.				
W	−12	−17	−21	− 6
N	9	31	16	41
3. Negro lawyers are as well prepared for the practice of law as are white lawyers.				
W	45	−10	−18	7
N	55	48	35	51
4. Where the litigation is between a Negro and a white, the white man or woman is favored to win in court without regard to the merits of the case.				
W	−24	37	38	34
N	34	39	26	48
5. Negro lawyers do not prepare their cases as well as white lawyers.				
W	−16	−41	−38	−48
N	−53	−55	−48	−59
6. Negro jurors are more easily swayed than white jurors.				
W	15	25	31	9
N	−29	−17	−20	−15
7. Negroes give too much irrelevant material in their answers to questions in court.				
W	24	30	31	23
N	1	24	23	24
8. Negroes should be represented on the staff of penal institutions in which the prison population contains Negroes.				
W	21	49	46	56
N	89	77	81	75
9. More severe sentences are meted out to Negroes than to whites for the same offense.				
W	− 8	46	49	34
N	64	54	51	22
10. A Negro represented by a white lawyer receives a lighter sentence than a Negro represented by a Negro lawyer.				
W	−43	24	28	13
N	26	26	31	24

Table 2—Continued

	Adults	College Students	Male College Students	Female College Students
11. A light complexioned Negro receives a more severe judgment by a white jury than a dark complexioned Negro.				
W	−58	−47	−38	−68
N	−44	−54	−53	−54
12. A Negro on a jury in the South is afraid to vote contrary to consensus of opinion of the white jurors.				
W	−21	55	55	53
N	62	35	28	39
13. A Negro litigant who is employed in a menial capacity by influential whites is favored to win in court over a Negro not so employed.				
W	−25	61	54	61
N	59	67	73	63
14. Judges have their minds made up before hearing a case when the litigation is between a Negro and a white.				
W	−76	−64	−54	−72
N	−21	− 3	− 6	− 1
15. Negro lawyers feel that Negro jurors are prejudiced in favor of the white side of the case.				
W	−41	−65	−62	−73
N	−24	−31	−34	−29
16. A Negro policeman arresting a white man cannot bring sufficient evidence against him to secure his conviction.				
W	−60	−58	−62	−41
N	−40	−56	−62	−51
17. Many cases of Negro conviction are found to be miscarriages of justice years afterwards.				
W	−11	17	26	− 6
N	48	60	58	61
18. Of several persons found flagrantly breaking the law, it is usually the Negro in the group who is arrested.				
W	0	40	33	53
N	66	54	45	61
19. Other things being equal, a white lawyer is favored to win in court over a Negro lawyer.				
W	7	71	75	60
N	54	43	36	48
20. A white woman's word in accusing a Negro is "proof positive" in court.				
W	10	20	23	13
N	46	55	51	57

Table 2—Continued

	Adults	College Students	Male College Students	Female College Students
21. Without provocation Negroes are beaten and otherwise maltreated by white policemen.				
W	5	−26	− 9	−54
N	66	53	52	54
22. White policemen who, without provocation, beat up Negroes are never convicted of their offense.				
W	− 3	−30	−19	−59
N	66	30	24	34
23. Negro lawyers will keep you in court the rest of your life.				
W	−25	−82	−82	−83
N	−76	−76	−78	−71
24. White juries are more prejudiced against the Negro on trial than white judges.				
W	−21	47	50	39
N	28	24	20	26
25. White lawyers make light of Negro lawyers in court.				
W	−29	26	31	14
N	− 1	9	20	2
26. White policemen do not cooperate with Negro policemen on the force.				
W	−32	−32	−31	−32
N	−51	−27	−34	−20
27. Judges are entirely free of racial prejudice.				
W	−26	−34	−36	−31
N	−72	−51	−61	−47
28. Negro lawyers lack the integrity of white lawyers.				
W	−40	−29	−24	−42
N	−93	−59	−59	−59
29. In the attitude of the court the Negro has no rights which the white man is bound to respect.				
W	−12	−73	−73	−71
N	−38	−20	−60	−22
30. Even federal courts countenance refusal of certain privileges of the court building to Negro lawyers.				
W	14	−12	−17	0
N	58	21	14	27
31. More white offenders are let off on insanity pleas than Negroes when both are accused of the same offense.				
W	38	60	72	48
N	60	72	72	71
32. White lawyers have more outside influence with the court than do Negro lawyers.				
W	49	89	91	83
N	65	82	88	78

Table 2—Continued

	Adults	College Students	Male College Students	Female College Students
33. In the South a white person must be quite a social outcast to be convicted of a crime against a Negro.				
W	26	38	40	35
N	75	66	56	73
34. The court believes a white lawyer is entitled to a larger fee than a Negro lawyer for the same piece of work.				
W	18	8	8	7
N	40	63	53	69
35. If a Negro is numbered among a group of suspects, he is usually the first to be grilled.				
W	41	60	62	54
N	73	71	68	73
36. Negro lawyers have few opportunities to practice other than routine office work.				
W	10	43	30	53
N	5	43	31	51
37. Negroes do not deserve the privilege of a court trial like other people.				
W	−67	−98	−96	−100
N	−100	−100	−100	−100
38. Negro lawyers do not take extra courses in law after receiving their law degree because they have little opportunity to use their knowledge.				
W	3	−28	−33	−17
N	−16	−22	−19	−24
39. Negro offenders under legal age are more often put into institutions with hardened criminals than white offenders of the same age.				
W	0	36	42	23
N	66	57	58	57
40. Many Negroes never bring suit against white persons, regardless of the amount of proof, because they feel that they are bound to lose in court.				
W	10	61	61	61
N	62	64	62	66
41. In cases of erroneous conviction of Negroes nothing is ever done to the person or persons who originally brought the false accusation.				
W	0	10	12	2
N	69	45	37	50
42. In criminal institutions Negro offenders are segregated into poorer quarters, given the most laborious tasks, and are in other ways maltreated.				
W	−10	14	15	14
N	74	50	45	54

Table 2—Continued

	Adults	College Students	Male College Students	Female College Students
43. In the South the sentencing of a Negro vagrant to a chain gang is usually a life-*sentence.*				
W	− 3	−65	−68	−57
N	12	13	−10	28
44. Negroes receive as much justice in courts as do whites of similar social status.				
W	−15	−11	−10	−13
N	−39	−37	−36	−38
45. A Negro lawyer in the North receives fairer recognition in court than a Negro lawyer in the South.				
W	− 5	77	78	75
N	78	85	88	83
46. Negroes would not know that they were treated unjustly if it were not for agitative organizations.				
W	−22	−33	−36	−27
N	−47	−49	−49	−43
47. White policemen accept bribes from Negroes only to have them arrested for bribery.				
W	−39	−67	−66	−70
N	−42	1	−19	15
48. White policemen accept bribes from whites and do not have them arrested.				
W	−41	27	35	3
N	25	40	30	46
49. Negro lawyers are shy about going up against white lawyers.				
W	−24	24	23	28
N	−54	−12	−26	− 2
50. Negroes in reality convict themselves in court due to ignorance of legal procedure.				
W	29	7	9	2
N	14	33	36	31
51. In democratic countries where everything depends on majority-vote courts are necessarily unfair to members of minority groups.				
W	−69	−23	−28	−10
N	−40	27	29	25
52. All cases of whites vs. Negroes or vice versa should be tried in Federal courts.				
W	− 8	−33	−31	−39
N	−13	38	42	36
53. A shabbily dressed, "hat-in-hand" Negro lawyer has more influence with the court than a well groomed, intelligent Negro lawyer.				
W	−64	−47	−42	−59
N	−46	−48	−40	−53

Table 2—Continued

	Adults	College Students	Male College Students	Female College Students
54. The victories Negroes win in court are left-handed victories.				
W	−24	−34	−32	−44
N	−22	−10	−16	−15
55. Negroes are not as financially able as whites to purchase justice in courts.				
W	33	65	68	56
N	73	76	77	75
56. It will take many more years before the Negro can have his constitutional rights granted him by courts.				
W	28	41	43	38
N	56	65	63	66

margins of questionnaires by white students as for examples: A white female student of the University of South Carolina with no court experience writes: "Some of these injustices against Negroes may be and probably are true, but they shouldn't be." A white male student from the same university writes: "Impossible to have an intelligent opinion on some of these questions. Negro lawyers aren't exactly numerous in the South." A white male student at the University of North Carolina and with court experience writes: "Sectional factors and differences between Northern and Southern courts are of such a nature that I cannot answer adequately. Northern and Southern courts are as different as night and day." A male white student of the University of Illinois says: "It is difficult to answer some of the questions unbiasedly because when they refer to the South they may be true, and false when they refer to the North." A male white student from the same university: "Negroes deserve as much justice as whites, but it is undoubtedly true that they receive less. It will take many years, if ever, for these social prejudices to be lived down." A white female student at the University of Florida writes: "I have no knowledge of the existence of a 'real' Negro lawyer."

A realization of the complexity of the situation, an appreciation of more than one side to the matter, a conflict between ideals and practice, lack of familiarity with the problem, little or no emotional involvement in the matter appear to be some of the factors making for mixed reactions so conspicuous in the whites.

In Table 2 are presented the 56 statements and the TF Indices of the various Negro and white groups for each statement. From Table 2 it is possible to make a direct comparison of any two groups in their reaction to a specific statement.

In Table 3 are presented the Pearson correlation coefficients of the 56 TF Indices for the various pairs of groups.

Table 3

Pearson Correlation Coefficients of the 56 TF Indices for the Various Pairs of Groups Arranged in Descending Order

Negro College Males and Negro College Females	.95
White College Males and White College Females	.94
All Negro College Students and All Negro Adults	.93
Negro College Males and White College Males	.87
Negro College Males and White College Females	.79
All Negro College Students and All White College Students	.78
Negro College Females and White College Males	.77
Negro College Females and White College Females	.76
All White College Students and All Negro Adults	.76
All Negro College Students and All White Adults	.72
All White College Students and All White Adults	.62
All Negro Adults and All White Adults	.60

Conclusions

The most significant conclusions which may be drawn from the results of this study appear to be as follows:

1. While Negro and white attitudes towards the 56 statements relative to the administration of justice as affecting Negroes correlate to a significant degree, this resemblance is much higher between Negro and white college students (.78) than it is between Negro and white adults (.60).

2. The correlation between attitudes is conspicuously higher between Negro college students and the Negro adults (.93) than between white college students and the white adults (.62).

3. While the resemblance in attitudes between the white college students and the Negro college students is high (.78), conspicuously higher is the resemblance between white college males and white college females (.94) and between Negro college males and Negro college females (.95).

4. The attitudes of the Negro college females and of the white college females (.76) correlate less than do the attitudes of any other two interracial groups among college students.

Received August 17, 1944.

Values Students Reported from the Study of Emotions

Key L. Barkley

Woman's College of The University of North Carolina

Both the friends and enemies of psychology have criticised it as an undergraduate course in the university on the grounds that it contributes very little of real value to the student. The opinion has been offered that perhaps the worst defect in the elementary course is that it is evasive in application, and that this weakness prevents the student from securing material which actually functions in his life. That is, the student may get a wealth of psychological fact, but receive little of psychological value.

Purpose

The purpose of the project reported here was not to discover simply what or how much the students *knew* about the facts, laws, and principles which characterize and govern human behavior, but to find out what values in any way or of any kind they believed they had secured through the study of a topic in psychology. The word "value" was interpreted to the students to mean any benefit, help, or gain they had received from the study, or any detriment, hurt, or loss they had suffered. In order to make the findings as definite as possible, to give the students a restricted topic on which to formulate judgments, and to increase the probable accuracy and dependability of the findings, the study was limited to the topic of "emotions."

Subjects

Two hundred twenty-six students of elementary psychology, who recently had completed their study of the topic of emotions, contributed their statements of values received. All of these subjects had the same readings assigned from two basic texts, and all of them had about the same number of lecture-discussion periods, namely six.

Procedure

Since there was no instrument available for the students to use in stating their judgments of values received from the study of emotions, the investigator had to devise one. To that end, forty-eight students were asked to write essays of about 200 words on the topic "The Values

I Have Received from the Study of Emotions." The experimenter read the essays and made a list of all the values the students said they had received. From the collection of statements, which were kept in the students' own words as far as possible, a check-list was made with the various stated values classified under nine arbitrary headings. The changes in statements when made were simply to shorten or to generalize the item.

The completed check-list was given to the students with the directions to check all the items which were statements of values they had received from the study of emotions, and to write in statements of values received which were not given in the list. It was made plain to all the students that their work with the check-list would have no relationship to their standing in the course in psychology which they were studying at the time.

Results

The number of people who checked the various items on the check-list was tabulated, and the per cent of the total of 226 subjects who checked each item was calculated. These percentages are shown in the first column of Table 1, which also presents the items in the check-list used in the experiment.

Certain findings are reasonably clear from an analysis of the data. The most obvious one is that over 80% of the students say they received definite and known benefits from the study of emotions. Only 17.3% of the students said that the study of emotions had had very little effect on them as persons (see item number 73). Moreover, some of those who made this statement went on to explain that they had received some advantage from the study, because it would help them in their professions such as social work.

A second clear finding is that a large number of people said they had received the values given in the check-list. The average number of persons out of the 226 who checked each item indicating a favorable value received was 128, or 56.6% of the group. It is obvious also that each person must have checked a considerable number of items. The average number of items checked by each person out of the total of sixty-four possible ones was 35.8. This means that most of the students believed they had received a large number of definite values from the study of emotions. It is probable, however, that no student received a separate and distinct value for each item checked, since there is considerable overlapping between some items.

The check-list was fairly comprehensive, and adequate to give the students an opportunity to indicate all the values they believed they had

Table 1

Showing check-list on which to indicate values received from the study of emotions, and the per cent of the 226 subjects who checked each item.

Directions: Put a check in the blank space before each statement of a value you have received from the study of emotions. If you have received a value for which there is no statement, write out a statement of the value in the place provided by a blank number under each heading.

I. General Values

Per Cent Who Checked Item	
74.4	1. It has given me a better conception of emotions; learned what emotions are and what they involve; gained insight into my own emotional life.
78.8	2. Learned when emotions occur, that emotions are the natural result of strong stimulation.
62.8	3. Learned to look at my own emotional problems objectively.
52.2	4. Learned that emotions are governed through the autonomic nervous system.
74.3	5. Study of emotions has been interesting and informative.
45.6	6. Learned about a factor that has greatly influenced student life.
71.7	7. Discovered that psychology has some practical values right now.
29.2	8. Improved my own general happiness through improved adjustment.
42.0	9. Established a firm background on which to base feelings and attitudes.
——	10.*

II. Development of Emotions

86.7	11. Learned that the aim in emotional development should be to direct and control rather than to inhibit emotions.
31.4	12. Learned that there is a native "core" of emotional behavior.
82.3	13. Learned that emotions are behavior, hence subject to change and development, especially through learning.
77.4	14. Learned that other people have a great influence upon our emotional development.
40.3	15. Discovered some good ways/means of emotional development.
83.6	16. Learned that emotions can be recognized, understood, and then corrected, or directed and controlled.
53.1	17. Learned what should be emphasized in the emotional education of children.
51.3	18. Learned that emotional re-education is possible; learned about emotional re-education, including the methods for removing undesirable emotional traits.
46.0	19. It has taught me which traits to avoid and which ones to attempt to attain.
35.4	20. Learned how emotions are differentiated.
54.9	21. Learned how emotions are established and developed.
45.1	22. Set up new goals; am now striving to emulate the standard of emotional stability set up in the course; I will try to meet the next set-back with firm resolution and not give way to emotion.
——	23.*

Table 1—Continued

Per Cent Who Checked Item	
	III. Role of Emotions in General Adjustment
74.3	24. *a.* Emotions as Motives. Learned that emotions serve as motives.
53.1	25. *b.* Emotions as Motives. This study motivated me to better efforts toward more wholesome emotional living.
66.8	26. Learned that emotions serve as facilitators and inhibitors of behavior in general.
83.6	27. Learned that a successful life must be founded on a firm emotional basis; a successful life is greatly dependent upon emotional stability and emotional maturity.
76.1	28. Helped me to see the important role they play in our lives.
——	29.*
	IV. Values Connected with the Experimental Study of Emotions
70.8	30. Learned that emotions cannot be judged from facial expressions.
64.6	31. Learned that the study of emotions is incomplete.
54.0	32. Awakened a new interest in the further study of emotions.
27.4	33. Learned how to measure emotions.
59.2	34. Discovered the difficulty of measuring emotions experimentally.
66.4	35. Made me more aware of emotions, viz., just what occurs in emotional action.
77.4	36. Learned how closely related are fear, joy, and rage in the effect on the body.
——	37.*
	V. Influence of the Study of Emotions upon Relationships with Others
65.9	38. Made student more tolerant of other people's conditions. I am now not as apt to criticize friends of mine.
50.9	39. Improved my adjustment to other people.
50.4	40. Developed better insight into the emotional life of other people; I am better able to help them.
38.5	41. Enabled me to aid another person to overcome an emotional depression.
78.3	42. I am able to look at others' reactions more objectively.
——	43.*
	VI. Values Connected with Emotional Maturity
62.8	44. Learned what constitutes emotional maturity, hence can work toward it now. I know what is expected of me as a college student. I realize the attributes possessed by the emotionally mature person.
30.5	45. It helped me to tie together all the loose ends in order to make more concrete my philosophy of life.
60.6	46. Aided student to recognize her limitations and powers. I have discovered where I fall short in being emotionally mature.
39.4	47. Discovered cause of poor adjustment to be emotional immaturity and instability, and learned what to do about them.

Table 1—Continued

Per Cent Who Checked Item	
72.1	48. Learned some of the causes of emotional immaturity.
18.6	49. Helped me to emancipate myself from my home.
79.6	50. I have seen the importance of being emotionally mature and stable.
53.5	51. Learned some of the necessary ways of solving children's problems. Learned how to help them become emotionally mature.
——	52.*
	VII. Values Connected with Emotional Stability
56.2	53. Learned what constitutes emotional stability, and how to work toward it. It is very helpful to know what makes a person emotionally stable and how to acquire these things.
40.7	54. Learned means of emotional control/restraint.
73.0	55. Learned that if a person will do something about a troublesome situation he will worry less.
32.7	56. I have gained fuller mastery of myself; attained greater emotional stability.
31.4	57. Learned how to overcome unfortunate moods.
81.9	58. Learned how important it is to your health and well-being to be emotionally stable.
——	59.*
	VIII. Special Individual Benefits
13.5	60. Helped me through an emotional crisis.
25.2	61. Aided in improving efficiency in study and work.
79.2	62. Learned that my emotional problems are not unique, but probably many others have them.
70.8	63. Learned that intense emotion may be harmful.
28.8	64. Lost queer notions, fear and apprehension regarding emotions.
23.9	65. Helped me to quit emotionally colored thinking.
64.2	66. Learned that most fears are unnecessary and handicapping.
40.3	67. Evolved the practice of being honest with myself as well as with others.
64.6	68. Study of emotions made me think and wonder about myself.
33.2	69. The study of emotions has made me more hopeful.
67.3	70. Learned that the facial expressions of others are not a safe guide to conduct.
50.0	71. Learned some steps to take in solving emotional problems and will be better prepared to meet emergency situations.
——	72.*
	IX. Negative or Questionable Values
17.3	73. The study of emotions has had very little effect on me as a person.
——	74.*

* Items numbered 10, 23, 29, 37, 43, 52, 59, 72, and 74 provided an opportunity to write in statements of values received not given in the questionnaire.

received from the study of emotions. Only four new items were listed by write-ins in the spaces provided for such additions. The remainder of the total of thirteen write-ins were criticisms of the course in psychology, or restatements of items already in the check-list.

All students did not secure significant values from the study of emotions. Seventeen per cent indicated that they had been helped very little by the study.

Some Conclusions and Criticisms

1. The method used in this study is in many ways subjective in nature, hence liable to errors which could not be checked on either as to their presence or degree. For example, many of the stated values which the students say they have received may be simply verbalizations not accompanied by real changes in the personalities and adjustments of the students. This weakness is indicated by the fact that some subjects checked a number of items as values received, but checked also the statement to the effect that the study of emotions had had very little effect on them as persons. But the fact still remains that more than 80% of the group indicated without reservation that they had received many positive values from the study of emotions.

2. The findings presented here represent the reactions of students in one university only where the first course in psychology is taught with a certain emphasis. In other universities where the first course is taught with a different emphasis, the responses of students might differ greatly from these.

3. This survey is a preliminary one, as it obviously must be, since the number of subjects is small. The emphasis upon the general benefits received from a course of study, rather than simply upon the knowledge of facts secured from it, is healthy, and the technique employed here appears to be a useful one. In this connection it should be pointed out that the values listed by the students were about equally divided between those which were strictly a matter of acquired knowledge and those which were of the nature of some new skill, changed viewpoint, better method of adjustment, etc. But the average per cent of the group who checked the knowledge values was 65, whereas the average per cent checking the other values was 48%.

4. Even though 80% of the students said they received many values from the study of emotions in elementary psychology, something should be done to reduce the percentage of students who say they get little or no benefit from it. In this University over three-fourths of the students who take the first course in psychology in normal times do not go on to

higher courses. The first course is often psychology's last chance to bring benefits into the lives of students through classroom instruction.

Received September 18, 1944.

References

1. English, H. B. Why students register for psychology. *J. appl. Psychol.*, 1928, 12, 242–244.
2. McGarvey, J. W. Interest in psychology as affected by the study of introductory course. *Psychol. Bull.*, 1938, 35, 668.
3. March, C. J. A student evaluation of course objectives in psychology. *J. genel. Psychol.*, 1942, 60, 381–384.
4. Schoen, Max. The elementary courses in psychology. *Amer. J. Psychol.*, 1926, 37, 593–599.
5. Tussing, Lyle. What students want from the elementary course in psychology. *J. appl. Psychol.*, 1938, 22, 282–287.
6. Wolfle, Dael. The first course in psychology. *Psychol. Bull.*, 1942, 39, 685–712.

Aircraft Recognition: I. The Relative Efficiency of Teaching Procedures *

Lester Luborsky

Duke University

Thus far the teaching of aircraft recognition has not been subjected to any reported experimentally-controlled analysis of learning product or comparison of teaching techniques.[1] The present Navy recognition program is based largely upon the groundwork of general principles arising from the research of Renshaw of Ohio State and his students, Schwarzbek (8), Knight (3), and others (1, 2, 5, 10). Their experiments led to such current emphases in recognition training as high speed presentation and perceptual learning of wholes rather than verbal learning of parts. Verbal learning of parts is emphasized in the WEFT system in which the characteristics of 4 main parts, the wings, engine, fuselage, and tail are memorized.

The standard instructional procedure—which has proved highly successful—used in most Navy recognition schools is somewhat as follows: When new planes are first introduced the shutter is set on "time" exposure—i.e., the shutter remains open and the airplane on the projection screen is visible until the plunger is pressed again—while good recognition features are pointed out by the instructor. Then, in the regular practice sessions which follow, the views are exposed for a short period, from 1″ to 1/50″, depending upon the particular school, and identification is made by the student. After each such identification, except during tests, the same view is reexposed for 3″–10″ and reidentified and discussed. It is thought best to use a large number of different views of each plane. The number of different views varies from 5 to 6, to a new view each time the plane is shown. The most usual rate of introducing new planes is 2 planes per session.[2]

---

* Part of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences of Duke University, 1945. Grateful acknowledgment is made of the assistance of two members of the Department of Psychology, Duke University: Dr. Karl Zener for sponsorship of the project, and to Dr. Sigmund Koch for valuable advice.

[1] Since completion of this experiment a summary of the research by Staff, Psychological Test Film Unit (9) has reported work on aspects of training procedures related to those dealt with in the present report.

[2] For valuable discussion of current training techniques and for access to classes the author is indebted to Lt. Comdr. H. L. Hamilton, Lt. W. C. Schwarzbek, and Lt. Comdr. R. H. Bruce of the U. S. Navy Pre-Flight School, Chapel Hill, North Carolina.

Renshaw and later workers developed the present program empirically on the basis of their past research findings. As a consequence, many differences in practice resulted, such as the variation in use of training exposures from 1/50″ at some schools to 1″ at others. Variability also occurs in the number of views presented and in the rate of teaching new planes.

It thus seemed desirable to test the relative effectiveness of several present differences in practice, and, in general, to subject a number of major variables of the training technique to experimental analysis. In brief, the object of the experiment was to determine, as exhaustively as possible, the effect of 4 basic variations of aircraft identification instructional procedure on the total learning behavior of equated classes of subjects. The design of the experiment was essentially:

1. The administration of a battery of tests calculated (*a*) to measure possible perceptual and learning factors associated with "aircraft identification ability" and (*b*) to aid in equating the groups for possible factors affecting trainability.

2. The actual training of 4 equated groups under different instructional conditions.

3. The administration of an extensive battery of post-training tests in order to facilitate more complete analysis of the learning product for the 4 groups.

The resulting data offer the opportunity for analysis of many of the determinants of aircraft recognition ability. The present report will be restricted to the major question of the relative efficiency of performance as a function of 4 training techniques. A future article will take up the role of other factors, including the span of apprehension, various measures of visual memory, and a fuller treatment of reaction time.

Procedure

Four equated groups, 8 Navy students in each (6 in Group I), were taught aircraft recognition in a standardized manner for 45 minutes on Monday, Wednesday, and Friday for a 6 week period in May-June 1944. For 3 of these groups one different experimental condition was varied.

The experimental groups were designed to compare the effects of two kinds of short exposure,[3] 1/50″ vs. 1″, both used in conjunction with approximately the same number of long exposures of the same stimuli (Group I vs. Group III);[4] the presentation of a large vs. a limited number of views (Group I vs. Group II); and the slow introduction of new planes *vs. a rapid introduction, followed by review* of confused planes (Group I vs. Group IV). In other respects an attempt was made to make the procedures comparable to those most frequently used in Navy recognition schools. For all groups in which tests were given at 1/50″ during training, the speed was attained by progressively reducing the exposure time during tests from one session to the next—starting at 1″ at the beginning of the second week and reaching 1/50″ by the last session of the third week. At each session three new planes were taught until, by the middle of the fourth week, 33 planes had been presented. By the end of the training period approximately 8 different views of each plane were used.

---

[3] For the purposes of simplicity, exposure time will be called 1/50″ and 1″, although the calibration values give slightly different results (see Apparatus, p. 389).

[4] The total number of 1/50″ exposures given to Group I was approximately 195 up to and including Test 9. An approximately equal number of exposures between 1/10″ and 1/50″ were given in tests before the 1/50″ speed was attained.

In Group II the exposure time and other variables were similar to Group I except that the number of views of each plane was limited to 3 diagram views. Group III was given 1″ exposures and Group I, 1/50″ exposures during tests. As will be indicated below in the description of training procedure, exposures for review and teaching were longer.

In Group IV an answer was sought to the question of the effect of teaching new stimulus material at almost twice the usual rate and then, after completing the syllabus, giving special instruction on the most frequently confused stimuli. As part of this special instruction, the most frequently confused planes were shown simultaneously on the screen and differentiation features stressed. When a large group of difficult homogeneous views were reviewed in this manner, as e.g., head-on-views of *single engine fighter planes, the students sketched these while learning.*

Training Procedure. Training, as described in the paragraph below, is composed of all the procedures which were directed toward the subjects' learning of the present aircraft recognition task. Training includes the frequent aircraft recognition tests and the review which followed these tests during which each plane was reexposed Training also includes the teaching of new planes and the review following this teaching, and the home study done by each subject. The paragraph below describes the training aspects of a typical session. In addition to this, training continues outside of class in home study. All groups were given the same general training procedures, with the exception of the experimental training variables mentioned above (exposure time on tests, number of views of each plane, and the rate of introducing new planes).

During the first 3 minutes required for dark adaptation, announcements were made of scores in the last test, etc. Then a test was given of planes taught up to that session in which subjects recorded the name of each plane on a specially lined mimeographed sheet. This required 10 to 18 minutes, depending upon the number of planes, which increased in successive tests as training was continued. The rate of presentation was 4 per minute. The experimenter, by watching the face of a large sweep-second hand clock, maintained a rather constant presentation tempo. All planes in this test were then reviewed for instructional purposes on time exposure. The time exposures were approximately 3″ to 6″. The subjects called out the name of each plane and were corrected when necessary. Subjects' questions were answered at this time to avoid making the review too mechanical. In the last 4 training sessions and in the "training-II"[5] sessions, the exposure time during this review was changed from about 3″–6″ to one or more 1/10″ exposures in order to give more practice with short exposures. For Group III the review exposure time was changed to 1″ at this time. The number of presentations in this review was determined by the number needed to equalize the total number of presentations for each group for the entire course, e.g., Group IV was given fewer since more planes were covered in each session. Approximately 3 minutes were then devoted to the teaching of each new plane. The head-on, plan, and side diagrams were viewed in succession, and good recognition features were emphasized by the instructor. After each view, reminders were given to attempt a visualization of the plane from all angles, to avoid irrelevant cues, and to make outline sketches while learning the plane. In the early part of the course, some suggestions for efficient study had been given and these were frequently reiterated. In addition, an interest item was mentioned for each plane, such as its use, or an example of its outstanding performance. The new planes were then reviewed on time exposure while the class called out the code number and name. Exposures were then speeded up to about 1/5″ in the latter part of this review. For Group III, however, no exposures less than 1″ were given.

Following the 4 week training period, for which the instructional procedures have been described above, was a 2 week test and "training-II" period. During these 2 weeks training was continued except that now all classes were given identical tests and trained to the conditions of Control Group I, i.e., all groups were given 1/50″ training, and picture as well as diagram views were included. In this 2 week period, 3 main training-II tests were given: Test 10, Test 11, and the Final Test. These tests will be described later in the section called "Tests of Recognition Performance during Training." Other tests were also given in other sessions of the training-II period, which will be referred to as "post-tests" in contradistinction to the "pre-tests." These post-tests were either repetitions of the pre-tests or tests of other aspects of the learning product.

To achieve uniformity in the use of study materials, all students were supplied with two standard aircraft recognition books (6, 7) and required to avoid any other material. A questionnaire on study methods filled in by the students near the termination of the training period shows that in general these books were the only material used.

---

[5] See paragraph following.

Apparatus

The 8 subjects in each group were seated about 5 feet from a projection screen. A constant seating position was maintained for each subject. The arrangement of seats was such that 4 subjects sat slightly out of the center line. This probably did not influence the pre- and post-tests or the rate of learning, since an analysis by seating position shows no differences related to seating. The experimenter operated a Baloptican, with an attachment supporting an iris-diaphragm shutter. This apparatus was placed immediately behind the students. Stimulus cards were fitted into a specially constructed postcard holder[6] which permitted ready changing of cards—in as short a time as 4 seconds. When the ready signal was given, the subjects fixated the center of the screen, and immediately after, the exposure was released by pressing the shutter-plunger.

Photographic calibration of the shutter was performed twice before and once after the experiment. The shutter speeds called 1/50″ and 1″ in this experiment gave calibration values of 1/40″ and 4/5″ respectively. Percentage changes from before to after the experiment were less than 6%.

Two sets of screen illumination readings were taken with a Macbeth Illuminometer. One set was taken with the room in semi-darkened condition and the other with the shutter opened on time exposure (with a blank card in the Baloptican). In the semi-darkened condition the screen reflected evenly 0.28 foot candles from a 15 Watt gooseneck lamp in the center of the room. The subjects recorded their responses to tests in this low light. The screen reflected evenly 2.55 foot candles of light when the shutter was open on time exposure. This was the illumination reflected from the screen when a plane was shown.

Twenty-five American and 11 British planes were selected for the syllabus because of their frequent use in combat.

Comparison of Group Scores

What is the effect of each of the experimental variables on the learning of each group? The complete data for training-I and training-II tests are given in Figure 1. The actual per cent correct scores for 5 of these tests which are especially important are given in Table 1A. The first of these, Average of Tests 1 and 2, may indicate the amount of the differences existing between groups before training. Test 9 indicates the effect of the experimental variables when tested under the conditions of these experimental variables. Test 10 is the most important test of all since it is given to all groups under identical conditions immediately after

---

[6] This postcard holder was constructed with a thin slot through which the stimulus card could be slipped in place for projection without the necessity of removing the postcard holder from its position beneath the Baloptican to change stimulus cards.

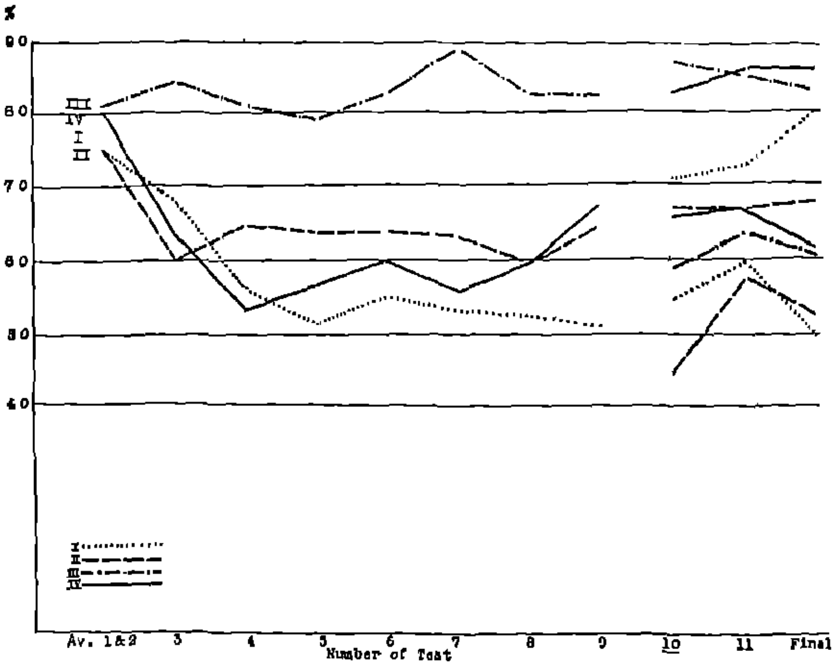


FIG. 1. Per cent correct scores on 9 training-I tests and 3 training-II tests. (See text.) The 4 upper score curves for training-II tests were given at 1″ exposure and the lower at 1/50.

completion of training-I and reveals most clearly the differential effects produced by the different training procedures. Test 11 and the Final Test are important in showing the consistency of the differences, but measure only residual effects of the training procedures. In the more detailed presentation to follow, the scores on Test 10 are italicized to emphasize their importance.

1. *Limitation in Number of Views: Group I vs. Group II.* A comparison of scores for Group I (the 1/50″ exposure group) and Group II (the limited-view group) indicates a small but consistent inferiority in Group II. Scores (per cents correct) on Average of Tests 1 and 2 and

Table 1A

Per Cent Correct Scores on Two Training-I and Three Training-II Tests

	Training-I Tests				Training-II Tests			
	Av. of Tests 1 and 2	Test 9	Test 10		Test 11		*Final Test*	
Exposure Time Group	1″	1/50″	1/50″	1″	1/50″	1″	1/50″	1″***
I	.75	.51	.55	.71	.60	.73	.50	.80
II	.75	.65	.45	.67	.58	.67	.53	.68
III	.81	.82*	.59	.87	.64	.85	.61	.83
IV	.80	.62	.66**	.83	.67	.86	.62	.86

* Group III was given 1″ exposure.

** Two weeks had elapsed between completion of syllabus and this test because of the accelerated schedule for introducing new planes.

*** This test may be somewhat less reliable than the others since scores are based on a test of only ten planes.

Table 1B

Per Cent Correct of Picture and Diagram Views *for each 1/50″ Training-II Test*

Group	Test 10		Test 11		Final Test	
	Pic.	*Diag.*	*Pic.*	*Diag.*	Pic.	*Diag.*
I	.50	.60	.55	.64	.48	.52
II	.28	.63	.52	.63	.48	.58
III	.52	.67	.65	.62	.58	.64
IV	.58	.74	.66	.68	.61	.63

Table 1C

Recognition Time for Planes *(recorded on the planes in* Test 10 in 1/60″ units)

Group	Recog. Time
I	69.7
II	56.9
III	54.3
IV	51.1

the three 1/50″ training-II tests are respectively as follows: Group I, 75, *55*, 60, and 50%. Group II, 75, *45*, 50 and 53%. The reliability coefficients of the differences between groups as indicated by *t* ratios are all greater than the 5% level of significance. (See Table 2.) For the 1″ training-II tests, however, scores for Group I are somewhat better than for Group II. The reliability coefficients of the differences are again above the 5% level, *but all favor Group I.*

Analysis of the percentages of picture vs. diagram views recognized correctly on Test 10, Test 11, and Final Tests (see Table 1B) reveals the source of Group II's low scores as due to poorer performance on picture as compared with diagram views. In Test 10, scores for diagram views are almost the same whereas for picture views the scores are 50% vs. 28% for Groups I and II respectively. Of course, the particular picture views used were completely new to both groups, the past difference in training being that Group I (and other groups) had had practice with other views and Group II had had only the three stereotyped diagram views.

2. *Exposure Time: Group I vs. Group III.* Group III (the 1″ exposure group) appears to be somewhat superior to Group I. On Average of Tests 1 and 2 and the three 1/50″ training-II tests, scores for Groups I and III are 75, *55*, 60 and 50% and 81, *59*, 64, and 61% respectively. The corresponding *t* ratios are statistically unreliable but all favor Group III.

Another aspect of the relative efficiency of performance of Groups I and III can be measured in terms of recognition time. This comparison should be important because proponents of the use of 1/50″ exposures in training may maintain that there are other advantages which do not emerge from comparisons of per cent correct scores, especially the important advantage of more rapid recognition. In Test 10, given immediately after completion of training, subjects were required to indicate recognition by lifting their forefinger from a response key in circuit with a standard timer. The name of the plane, given verbally by the subject, was recorded by an assistant. The results in Table 1C show conclusively that an advantage for Group I in recognition speed did not exist. Furthermore, accuracy and speed of recognition in Test 10 for all subjects proved to be uncorrelated ($r = -.16$).

3. *Rate of Introduction of New Planes: Group I vs. Group IV.* Group IV (the rapid-presentation-plus-review group) is superior to Group I. Scores on Average of Tests 1 and 2 and the three 1/50″ training-II tests were 80, *66*, 67, and 62%, as against corresponding values of 75, *55*, 60, and 50% for Group I. The *t* ratios for the differences between Groups I and IV on training-II tests are 1.83, 1.31, and 2.29. Although only the last of these is significant at the 5% level, the direction of all of them favors Group IV.

A second, and obviously related aspect of the results concerns the intercomparison of all groups. Groups III and IV give the best results of all 4 groups. After completion of the 4 weeks training period, it was found that Group III did as well as both Groups I and II at 1/50″ exposures (Test 10) *even though no previous training on planes at 1/50″ had been given.* Furthermore, Group III did somewhat better on this test than any other group with 1″ exposures. Group IV seems to be slightly better on the Final Test at both 1/50″ and 1″ than any other group. The *t* ratios for the differences between Groups II and III on the three 1/50″ training-II tests are 1.41, 0.83, and 1.25, and for the Groups II and IV they are 2.49, 1.59, and 1.95.

Final evaluation of the differences in this section depends upon the reliability of the differences and the comparability of groups which are discussed below.

Reliability of Differences

How much confidence can be placed in the obtained differences? Since the number of subjects in each group is small and the absolute values of many of the differences not particularly great careful consideration must be given this question.

Table 2

Reliability of Differences Between all Groups on Average of Tests 1 and 2 and Training-II Tests (*t* ratios) *

Group	Av. Tests 1 and 2	Test 10		Test 11		Final Test	
	1″	1/50″	1″	1/50″	1″	1/50″	1″
I II	0.00	1.20	0.33	0.23	0.80	0.46	1.23
I III	0.87	0.51	1.60	0.58	1.90	1.62	0.31
I IV	0.73	1.83	1.10	1.31	2.14	$\underline{2.29}$	0.82
II III	1.01	1.41	$\underline{3.04}$	0.83	$\underline{2.49}$	1.25	1.50
II IV	0.79	$\underline{2.49}$	$\underline{2.95}$	1.57	$\underline{2.66}$	1.95	2.12
III IV	0.14	0.80	0.03	0.82	0.28	0.20	0.53

* If underlined the difference is within the 5% level of significance, i.e., greater than 2.17 for those differences involving Group I and greater than 2.14 for others.

Table 2 presents the *t* ratios for the differences emerging from intercomparison of the performance of all groups on the Average of Tests 1 and 2 and the 3 crucial training-II tests for 1/50″ and 1″ exposures.

Although these reliability measures for the most part are not statistically significant, examination of Figure 1 will reveal that the results show a consistent direction. It is this consistency of direction which suggests that the main differences, i.e., those between the first 2 and the second 2 groups, may be real ones.

Comparability of Groups

More important than statistical reliability in the case of such small groups is the question of the rigor of equation of all variables other than the experimental ones. The groups were relatively satisfactorily equated before training for 2 factors with evident importance for trainability—intelligence and previous knowledge of planes. The results of this

Table 3

Average Scores for Each Group on Equation Tests
And Other Tests Which Could Have Affected the Trainability of the Groups

Group	A.C.E.	Previous Knowledge	Grades—Academic Work	Study Time—Aircraft Recognition		Acuity		Interest			Reaction Time	
	Total Scores	Planes Correct	Point Equivalents	Minutes		Letters Read		Planes Incorrect		% Imp.	Light	Objects
				2nd wk.	3rd wk.	Pre	Post	Before	After		1/60″ Units	
I	108.16	8.50	6.9	73	67	30.00	29.33	14.83	5.66	64	8.6	17.6
II	114.25	6.30	6.6	92	92	27.87	29.50	14.25	7.25	53	8.6	13.3
III	114.00	8.75	7.5	69	63	26.87	30.12	13.37	4.62	65	9.8	18.4
IV	105.50	9.00	6.7	51	51	27.12	29.63	10.25	4.75	52	8.6	17.4

Group	Span of Apprehension		Memory—Location	Memory—Complex Figures Part 1	Memory—Complex Figures Part 2	Memory—Complex Figures, 5″	Memory—Airplane Recognition Aspects
	Amount Above or Below	Number Incorrect	Number Incorrect Blocks	Number Errors	Number Errors	Number Errors	Number Errors
I	16.16	7.83	58.00	5.30	27.60	5.16	36.50
II	10.75	6.87	51.40	4.40	24.00	4.50	39.80
III	17.87	8.37	50.60	4.10	22.00	3.63	35.80
IV	14.12	7.90	49.60	4.30	21.00	4.03	39.80

equation will be presented in Table 3 together with the results of tests of the other factors which might have been responsible for differences in the trainability of the groups. In the following list of these factors, the name of the measure alone gives some suggestion as to the nature of the measure. A complete description of each will be given in another report (4). The list is: 1. Intelligence (A. C. E.); 2. Previous Knowledge of Planes; 3. *Grades in Academic Work for Previous Semester;* 4. *Home* Study Time for Aircraft Recognition; 5. Acuity (Snellen Symbol E Chart); 6. Interest I; 7. Interest II (Improvement); 8. Reaction Time (Light Flash; Common Objects; Airplanes); 9. Span of Apprehension, 1/50″; 10. Memory-Location, 1/50″; 11. Memory-Complex Figures, Part 1 1/50″; 12. Memory-Complex Figures, Part 2 1/50″; 13. Memory-Complex Figures, 5″; and 14. Memory-Aircraft Recognition Aspects, 1/50″.

Although the differences between groups on the above tests (see Table 2) appear small enough to disregard if taken singly, Group I and to a lesser extent Group II, are lower on several tests (Memory-Complex Figures 5″ and Memory-Complex Figures, Part I, 1/50″) which have been found (4) to be important in determining final level of performance. Since the effect of these differences is not determinable, the differences between groups in aircraft recognition tests must be somewhat larger to be meaningful than they otherwise would have to be. This will primarily affect the conclusions based upon Group I vs. Group III.

Other differences not covered by the above measures may have existed. For example, the differences in group scores obtained on Average of Tests 1 and 2, ranging from 0% to 6%, possibly were not produced by the experimental variables in training procedure, although these variables may have slightly affected the scores of Group IV and Group II. This may be a crucial criticism. However, a reanalysis of the data on the basis of a re-equation of groups shows that it does not materially affect the results. The re-equation was carried out in the following manner: Average of Tests 1 and 2 scores of 4 subjects, one from each group, were selected on the basis of how nearly they were alike. It was possible to find one low, high, low-middle, and high-middle score which was almost the same in each group. The means of these 4 sets *of scores, one for each group, did not differ more than 0.5%. These* re-equated group-means on Average of Tests 1 and 2 were then compared with the mean of the same 4 subjects on Test 10, Test 11, and the Final Test. Virtually the same relative differences among the scores were obtained as with the original data. The re-equated Group IV, however, was slightly higher on the Final Test.

Discussion

A number of suggestions for optimum training procedure may be immediately inferred from our results:

1. Group II procedure seems to be the least efficient. Therefore teaching materials which are restricted to only 3 diagram views of each plane, as are those of many spotters' courses, are disadvantageous for proper learning, and cause difficulty in recognizing other, more life-like views.[7]

2. A training procedure including tests at exposures varying from 1″ to 1/50″ has no evident advantages over a procedure having no exposures shorter than 1″ (Group I vs. Group III). The slight apparent superiority of Group III on aircraft recognition tests is difficult to evaluate because of the small initial superiority of this group on Average of Tests 1 and 2 and on some of the pre-tests which are correlated highly with final level of performance. There are three possible interpretations to the failure of Group I to show any superiority over Group III on Test 10 which constituted the first experience of this group with 1/50″ exposures. There may, in fact, actually be no special skill for seeing under short exposure conditions. The development of such a skill may not readily occur in the absence of conditions permitting operation of the law of effect, that is, more immediate identification of the plane verbally or by re-exposure for a longer time. Or lastly, a longer period of training with 1/50″ exposures, as is usually found in recognition programs, may be necessary.

The data presented in the next article (4) would enable a more complete equation of groups than was possible before this experiment and would open the way to a definite answer to the problem.

Since the speeds tested in the present experiment are within the range considered short exposure, they offer no direct implications as to the relative advantages of short vs. long exposures. (Long exposures are here considered as 3 to 10 seconds.) Possibly some combination of long and short exposures may be more efficient than predominant use of either. This technique is employed in some recognition schools. In addition, more consistent and definitive use was made of the 1″ and 1/50″ exposures than is usual in current practice. Consequently, this afforded clearer insight into the effects of these procedures.

3. New planes can be learned at almost twice the usual rate with no impairment in efficiency, in comparable length of time to the other teaching procedures, by the rapid-presentation-plus-review procedure.

---

[7] In addition, there seems to be a greater tendency to use inadequate and trick recognition features. At the end of the course students were asked to confess any trick cues they had been using. Group II seemed to excel in their use.

There is strongly suggestive evidence that it is the most effective of the procedures used. Its superiority may be the result of a longer period of time during which all planes have been seen and a longer period of time for stressing those planes which present the most difficult differentiation problems.

Since the procedures for Group III and IV were found to be most efficient, a combination of both might give greater efficiency than either one separately. This is one interesting possibility for further research.

Summary and Conclusions

Four equated groups, of 8 pre-aviation (V-5) students each, were taught aircraft recognition in a standardized manner. The major experimental variable in each group was as follows: Group I, 1/50″ exposure time; Group II, only three views of each plane; Group III, 1″ exposure time; and Group IV, presentation of the entire syllabus in almost half the usual time, followed by a review emphasizing confused planes.

The following conclusions were obtained.

1. The use of 1/50″ exposures as part of training in which an approximately equal number of longer exposures are given has no ascertainable advantages over 1″ exposures similarly given with longer exposures.
2. Restriction of teaching materials to only three views of each plane results in learning which generalizes poorly to views of planes other than those taught and is, in this sense, inefficient.
3. Rapid teaching followed by review of confused planes is probably the most efficient of the procedures tested.

These findings might be useful for incorporation in actual training courses.

The battery of tests which was found (4) to be important for determining the final level of performance will now make possible a more complete equation of groups and therefore future experiments which can yield more definite answers to problems such as the above.

Received October 18, 1944.

References

1. Banner, A. *The effect of practice on the perception and memorization of digits presented in single exposure.* Ph.D. thesis, Ohio State University, 1935.
2. Bennett, S. *The visual perception of English words of various lengths in tachistoscopic exposures of 3 milliseconds.* M.A. thesis, Ohio State University, 1940.

3. Knight, O. D. *The role of the figure-ground relation in perceiving and memorizing visual forms.* Ph.D. thesis, Ohio State University, 1936.
4. Luborsky, L. Aircraft recognition: II. A study of prognostic tests. *J. appl. Psychol.* (in press).
5. McIntyre, S. C. *The role of summation and of some other variations of impression in the perception and memory for visual forms.* M.A. thesis, Ohio State University, 1939.
6. Pitkin, W., Jr. *What's that plane?* (3rd Ed. rev.) Washington, D. C., and New York: *Infantry Journal*, Penguin Books, 1943.
7. Saville-Sneath, R. A. *Aircraft recognition.* (3rd Ed.) Washington, D. C., and New York: *Infantry Journal*, Penguin Books, 1943.
8. Schwarzbek, W. C. *Some factors which influence the impression and immediate reproduction of digits.* Ph.D. thesis, Ohio State University, 1935.
9. Staff, Psychological Test Film Unit. History, organization, and research activities, Psychological Test Film Unit, Army Air Forces. *Psychol. Bull.*, 1944, 41, 457–468.
10. Steckle, L. C. The relative efficiency of single versus multiple exposures in the rapid memorization of visual forms. Denison University Bulletin, *J. Sci. Laboratories*, 1940, 35, 1–31.

Magazine vs. Personal Interview Votes in the Consumer Jury Advertising Test

Lester Guest

The Pennsylvania State College

For many years the consumer jury or opinion method of testing the effectiveness of advertisements has been a useful device for estimating the relative effectiveness of advertisements before they appear. In essence, respondents are usually personally interviewed and asked which of several advertisements interests them the most or would most likely influence them to buy.

The completeness of the make-up of the advertisements at the time of the test, the number of advertisements that can be judged at one time, and the form of the question posed the respondent have all been subjected to research (1) (3).[1] It has been conceded that, as in all sampling studies, the sample should be representative of the purchasers or potential purchasers of the product in question. The final answers to some of the other problems await some sort of definitive criterion of effectiveness. However, assuming that respondents' stated preferences are automatically valid for the degree to which they would be likely to *read* an advertisement, some questions can be tentatively answered.

In regard to the form of the question, the consensus seems to indicate that respondents should be asked to select the advertisement which appeals to them the most, or that interests them the most, or the advertisement that would most likely lead them to buy, rather than asking them to select the *best* advertisement (2, p. 369) (3, p. 124). (There are those who say that the wording makes little, if any, difference (1, p. 18).) The reason for this distinction in question wording arises from the belief that the interviewee should be asked to react to the advertisement *as a consumer* and not as a *critic* of advertising. From the current point of view, there is no *best* advertisement apart from the reader's own preference. From a practical point of view, it makes no difference whether the advertising expert believes that a certain advertisement has the best headline, general layout, and illustration, if the public concerned dislikes that advertisement and prefers another. Therefore, the question asked the respondent should not lead the respondent to believe that he is matching wits with experts, but should let him feel free to react as he *usually* acts, not as he *should* act.

---

[1] Unpublished data from several sources also give information on these points.

With respect to sampling problems, it has been shown that coupon or mail responses are not usually representative of the general population or of a specific mailed sampling. This is sometimes due to dilution or inflation caused by the habitual coupon clipper, or even more disconcerting, the inertia of the temporarily disinterested individual or the over-zealous individual who has a special interest in the problem at hand. In addition, Link (3) points out that mail inquiries allow respondents too much time for *critical* evaluation which may lead to distortion.

The author recently had an opportunity to study the following questions in respect to consumer jury responses: (1) what effect does the form of the question have upon preference for an advertisement, and (2) what differences, if any, result from a comparison of magazine ballots vs. personal interview ballots?

Procedure

The advertisements in question were printed in the interests of a well known, nationally distributed drug product. The original advertisement in reality consisted of two advertisements placed side by side on a full page of a national magazine. The headline of the complete advertisement asked the reader to indicate which ad he or she voted for, with a small subscript challenging the reader to "match wits with the experts." The two advertisements then appeared, followed by a short bit of copy and a coupon. In return for the reader's reply, he was to receive a free sample of the product. These two advertisements will be referred to as A and B. Advertisement B subsequently appeared singly in three other national magazines, but Advertisement A never appeared elsewhere.

The double advertisement asked the reader which advertisement was the *better*. In order that coupon replies and personal interview replies be comparable it was imperative that the personal interview question also ask which advertisement was better. However, it was thought desirable to check the influence of wording the question in this fashion with wording believed to be more appropriate, i.e. "most interesting." Therefore, the questionnaire was constructed with this aim in view.

Two forms of the questionnaire were constructed, each asking for data concerned with usage of the product, recognition of the advertisements, and readership of the magazine in which the advertisements originally appeared. In addition, each person was asked which advertisement *interested* him more and which advertisement he thought was *better*. On Form A of the questionnaire, the "interest" question preceded the "better" question, and on Form B the order of these two questions was reversed. Other than this, the two forms were identical. Additional data were secured bearing upon the respondent's age, sex, and economic status.

Proofs of the original advertisement were obtained and by appropriate cutting were made into two separate advertisements, each as they might appear singly. These were presented to the respondent simultaneously, but the right-left position was systematically varied to avoid any time-order error. In the original magazine appearance this was of course impossible.

A total of 304 interviews were conducted by experienced interviewers. These were done about 2 months after the appearance of the original advertisement and were stratified into the conventional A, B, C, and D economic groups. Half of the interviews were done with each form of the questionnaire, and half were done with each sex. No attempt was made to control the age distribution, although the approximate age was noted for each interviewee. Geographically, the interviews were done in cities and small towns distributed along the eastern coast.

Several internal checks of the data indicate that the sample was reasonably representative. For example, although no effort was made to interview the same number of men and women with each *form* of the questionnaire, about one-half men and one-half women were found to have been interviewed with each form. Similarly, although economic status for the whole 304 interviews was stratified, no attempt was made to stratify *within* one questionnaire form. However, approximately the correct proportions were maintained by form of the questionnaire. Finally, about the same number of users of *some brand* of the product were found to have been interviewed with each form of the questionnaire.

Results

The data presented in Tables 1 and 2 refer to the material collected from all interviews irrespective of the form of the questionnaire used. Cases where respondents refused to choose between the two advertisements have been eliminated throughout to facilitate comparisons. These constituted only 10% to 12% of the cases and their elimination did not materially change the results. The category "interesting" always will refer to the question phrased using that word, and the category "better" will refer to the question using the word *better*. Table 1 indicates that in all comparisons Ad A was judged superior regardless of the question asked or the way the data were collected. A's superiority

Table 1

Per Cent for each Advertisement in Total Interview Group and Coupon Group

	Ad A	Ad B	*N*
Personal Interview			
Interesting	53%	47%	266
Better	58%	42%	274
Coupon			
Better	59%	41%	239

Table 2

Per Cent for each Advertisement According to Economic Status, Sex, and Use of some Brand of the Product *

		More Interesting			*Better*		
		Ad A	*Ad B*	*N*	*Ad A*	*Ad B*	*N*
Economic Status	A	59%	41%	29	55%	45%	29
	B	51%	49%	53	60%	40%	57
	C	52%	48%	104	57%	43%	107
	D	54%	46%	80	57%	43%	81
Sex	Male	55%	45%	125	60%	40%	131
	Female	51%	49%	140	55%	45%	142
Use of *some* brand of the product	Users	50%	50%	103	54%	46%	105
	Non-users	55%	45%	163	60%	40%	169

* Only 3% of the total group of respondents used the *brand* of the product advertised and therefore this group was considered too small for any statistical analysis.

increases when "better" was asked in the personal interview and agrees closely with coupon returns in this case.

In no case, however, did the critical ratios between these differences reach 3. The difference of 18% in favor of Ad A in terms of coupon returns yields a critical ratio of 2.83, or over 99 chances in 100 that the difference is not the result of sampling errors. Likewise, the difference of 16 in favor of Ad A from the personal interview returns when "better" was asked gives 99 chances in 100 that the difference is not the result of sampling errors.

When the data are fractionated for economic status, sex, and use of *some* brand of the product, the base N's become too small to yield any statistically significant differences. However, such breakdowns can give *suggestions* as to the possible groups from which coupons were returned in order that Ad A receive 59% of the votes. Therefore, these data are presented in Table 2.

Some of these percentages agree quite closely with coupon returns, for example, the percentage of men picking Ad A as better, the non-users picking Ad A as better, and the B economic group selecting Ad A as better. From this, it looks as if coupons might have been returned with a greater frequency from men, the B economic group, and non-users of the product, otherwise coupon returns would not agree as well with personal interview data. As a matter of fact, it is known that coupons were returned with men's signatures more frequently than with women's signatures when the coupon returns were broken down for sex, 60% of men prefer Ad A and 55% of women prefer the same advertisement. This agrees perfectly with the interview data for sex on the same form of the question. The coupons did not contain data allowing other fractionations.

It will be recalled that on one form of the questionnaire, "interest" was asked first followed by "better" and that this procedure was reversed on the other form. The previous tables present the data grouped by question form but summating responses irrespective of the order of the questions. To make the most legitimate comparison with coupon returns it is necessary to consider only responses to the question asking which advertisement is better on the personal interview questionnaire, unbiased by a previous question asking which advertisement is more interesting. To do this, responses were tabulated separately for Form B of the questionnaire and only for the question asking for the "better" advertisement. This question appeared first on this form. Table 3 gives these results.

Table 3

Per Cent for each Advertisement According to Economic Status, Sex, and Use of the Product for the "Better" Question Only and When it Appeared First (Form B)

		Total Groups		
		Ad A	Ad B	*N*
Coupon (Better)		59%	41%	239
Personal Interview (Better)		53%	47%	135
		Fractionations for Personal Interview		
		Ad A	Ad B	*N*
Economic Status	A	50%	50%	14
	B	60%	40%	30
	C	58%	42%	53
	D	39%	61%	38
Sex	Male	55%	45%	64
	Female	51%	49%	71
Use of *some* brand of the product	Users	51%	49%	53
	Non-users	54%	46%	82

Here again, the N's are much too small to yield any statistically significant differences but the results shown here suggest that coupons probably were returned most heavily from the B economic group in order that 59% of magazine voters could have selected Ad A. On the whole, it is probable that coupon responses overestimate the "superiority" of Ad A even in terms of "betterness." (The critical ratio of the difference of 18% in favor of Ad A yielded by magazine votes is 2.83 whereas the critical ratio of the difference of 6% in favor of Ad A yielded by personal interview results is .70.)

Another factor that could have led to the large majority of coupon responders selecting Ad A was the fact that Ad A always appeared on the left position in the magazine whereas the right-left position was alternated in the personal interview. Other studies have shown the influence of time-order error and it is conceivable that it operated in this instance (1).

It is obvious that differences arise not only from the form of the question but also from the effect early questions in the questionnaire have upon other subsequent questions. Table 4 presents material upon this factor.

Table 4

Influence of the Order of Questions upon Preference for an Advertisement

	Form A				Form B		
	A	B	*N*		A	B	*N*
Interest (1st)	55%	45%	132	Better (1st)	53%	47%	135
Better (2nd)	63%	37%	139	Interest (2nd)	51%	49%	134

The material in this table shows again that Ad A is superior in every comparison, but the *degree* of superiority varies depending upon the context of the question. It seems that from a pure interest point of view, Ad A is picked by 10% more people than Ad B, but when followed by a question referring to "better," Ad A's superiority jumps to 26%. (The critical ratio of the latter difference is 3.17. In all other comparisons in Table 4 the critical ratio approached zero.) On the other hand, if "better" is asked first, followed by "interest," the change is only from 6% to 2% in favor of Ad A. It looks here as if, once a person has said that an advertisement is better, there is less likelihood of his turning about-face and choosing another advertisement as more interesting, but if he has picked an advertisement as more interesting, then is confronted with a question inferring that there is such a thing as a "better" advertisement, he may be more prone to change his choice. The data presented in Table 5 indicate that the large majority of people do *not* change their original selection, but that those who *do* change materially alter results for the total group.

Out of 152 answers to Form A, 107 made no change, and out of 152 answers to Form B, 109 persons made no change. Therefore, the appearance of either question before the other is not especially conducive to change of choice. However, considering only those that changed preference, when better precedes interest, the changes were about evenly split, whereas, when interest precedes better, twice as many persons change to Ad A as change to Ad B. (The critical ratio in the latter instance with *only 45 cases* is 2.37.) Evidently, there are some factors in Ad A *that respondents who originally select* B as interesting feel experts agree upon as better. What those factors are is not obvious from the present analysis.

Table 5

Major Changes in Choice of Advertisement by Form of the Questionnaire

Form A (Interest—better)			Form B (Better—interest)		
Changes to B when asked "better"	Changes to A when asked "better"	*N*	Changes to B when asked "interest"	Changes to A when asked "interest"	*N*
33%	60%	45	49%	51%	43

An interesting commentary concerning the whole study is that although Ad A always gathered the majority of the votes (sometimes small), Ad A was never published as a separate advertisement, whereas Ad B ran separately three times. This supports the contention that the copy writer, although he may be an excellent judge of an advertisement *for himself* or a *small* number of people *like him, may not estimate in* advance which of several advertisements will please the *majority* of the reading public with which he is really *most* concerned (3).

Conclusions

Any conclusions drawn must of necessity be interpreted in the light of the fact that, due to the relatively small number of cases interviewed, most of the differences do not yield critical ratios of 3 or more. However, in several of the more pertinent comparisons, the chances are greater *than 90 in 100 that the obtained differences* are not due to sampling errors. Remembering the limitations of the study, the following conclusions may be drawn.

1. A comparison of the results of a consumer jury test carried out by magazine votes with one carried on by personal interview indicates that the two give different results. Magazine returns indicated 59% for Ad A whereas the most legitimate comparison figure for the personal interview group was 53% for Ad A. The critical ratio of the difference is 1.12, or 87 chances in 100 that the difference cannot be explained in terms of sampling errors. At least one other study has shown a similar discrepancy (3, p. 119).

2. The differences found between magazine and personal interview results may be a result of unrepresentative sampling of coupon responses. It appears that the middle economic groups' responses more nearly approximate coupon returns than other groupings. The magazine in which the original advertisement appeared would tend to be read by the top economic groups more than the lower groups.

3. There is a possibility that a time-order error may be introduced in magazine voting and that this can be balanced out by a properly controlled personal interview study.

4. The form of the question has an important bearing upon results obtained. Different responses are obtained when a person is asked to select the better of two advertisements than when he is asked to choose the one that interests him the most. It is likely that the latter form will give a truer picture of him as a *consumer* than the former form.

5. For some people, the answers to a question will be unduly influenced by preceding questions and answers.

Received September 5, 1944.

References

1. Advertising Research Foundation. *Copy testing.* N. Y.: Ronald Press, 1939. Ch. 1.
2. Jenkins, John G. *Psychology in business and industry.* N. Y.: Wiley, 1935.
3. Link, H. C. *The new psychology of selling and advertising.* N. Y.: Macmillan, 1938. Ch. 6.

Book Reviews

Woolpert, E. D. [Ed.] *Municipal Personnel Administration.* (3rd edition.) Chicago: International City Managers' Assoc., 1942. Pp. xii + 429. $7.50.

Municipalities and other jurisdictions entertaining the idea of installing a personnel system, contemplating the institution of an in-service training program or seeking a comprehensive, realistic guide to workable solutions of personnel problems should consult *Municipal Personnel Administration.* This book is published by the International City Managers' Association as the ninth volume of its series on municipal administration.

While public officials are inclined to be interested in specific problems, rather than general practices and trends, they should be ever mindful of the need for perspective and long-range policies and programs. Accordingly, this volume consists of more than expedient devices for the practitioner. The first chapter on the personnel problem does just this by providing a broad background and a setting for the more specific discussions of personnel problems which encompass the remainder of the volume.

The need for a rational organization of personnel activities is the first specific personnel problem considered. It is indicated that the dimensions and complexity of this problem require positive, concerted effort for effective solution. A basic requirement for dealing adequately with this personnel problem is to have defined clearly the responsibility for the administration of such activities. This personnel problem exists even in the smaller cities where it is not always feasible to have a separate personnel agency. However, it is pointed out that these cities must recognize regardless of working force the existence of such problems as classification, rates of pay, recruitment, hours of work, attendance, and leaves of absence. The text outlines several ways whereby these municipalities can meet their personnel problems. The dominant theme is that variations in structural patterns of personnel organization should not be taken seriously as long as they result from the application of sound general principles to different local situations. The basic problem of organization is solved when the work is analyzed and the methods and resources available are considered and evaluated.

In the ensuing several chapters where the principal phases of personnel administration are considered, an attempt has been made to outline standards of administration as guideposts for individual personnel agencies. In the chapter on position-classification, it is shown that information relative to the duties and responsibilities attached to the various individual positions within a given service is of primary importance in the development of a personnel program. Accordingly a position-classification plan designed to secure and utilize these facts should be one of the first aims in preparing and administering a personnel program. After stating the nature and objectives of position-classification, serious attention is given to the development of a classification plan, the problems entailed in introducing the plan, and the basic requisite of continuous administration of the plan.

In a concise, yet complete, chapter on salary and wage administration, it is shown that the intangible advantages of public employment about balance the disadvantages. A workable outline is presented for those seeking to attain the principal objective of a pay plan—equal pay for equal work.

In Chapter 5 the reader finds an objective discussion of problems involved in the recruitment and selection of qualified personnel for public service. After indicating the relation of recruitment to classification, there is presented the various forms and methods of examining or testing applicants in selecting employees with a capacity to learn. In logical sequence there follows a chapter on employee training in which there is found most of the elements entailed in the organization for and methods of training. The point is made that the appropriate method of training should be determined pragmatically, on the basis of utility and applicability. In Chapter 7, considerable attention is given to the methods of appraising, through promotional test, the capacity of employees for promotion. Personnel administrators are reminded that there is an inverse relationship between the need for specific knowledge, the factor which a well-constructed examination can measure reliably, and the level in the organizational hierarchy which an employee occupies. The higher one goes in the organizational ladder the greater should be the emphasis upon capacity for administration.

The problem of reports of performance is treated in an unbiased manner. The evaluation of employees is regarded as an essential part of administration since it is quite common to find disparity between ability and performance. In any balanced personnel program determination of knowledge, skills, and aptitudes should be supplemented by measures of employee performance. Various types of rating systems are described and an effort made to evaluate the degree of success found with each. The analytic checklist type seems to be favored because it tends to eliminate the weakness of evaluation by the rating officer prevalent in the graphic rating scales.

What has been presented to this point is regarded as the skeleton of an effective personnel program. The chapter on morale and conditions of employment includes other elements which enter into such a program enabling it to give maximum service to the community. The quality of supervision within an organization is presented as the most important factor in building and maintaining high morale. Closely related to the question of morale is the complex matter of discipline. While favoring some form of disciplinary process, similar to the employee self-discipline plan inaugurated in the refuse collection division of the Los Angeles city government, responsible officials are cautioned against copying without discrimination this or any other plan as a general model. Since self-discipline can best be developed and maintained through the organization of employees themselves, the following chapter on employee relations delves into the many ramifications of the personal equation in administration, particularly as they relate to the development of a municipal employee relation policy. As the final element of a comprehensive personnel program, Chapter 12 goes into the ramifications of a properly planned and administered retirement system. While the personnel administrator does not usually administer the retirement system, it is advised that he participate in the planning phases because it plays an important part in the development and maintenance of effective government.

After covering the broad background from recruitment to retirement, Chapter 13, entitled "Special Administrative Problems," presents several aspects or phases of personnel administration which have to do with the overall administration of a personnel program, rather than with the program *per se*. They include such items as personnel rules and regulations, personnel forms and records, research, measurement, and public relations.

Practitioners, namely administrators, personnel psychologists, and heads of operating departments and agencies, will find the book extremely valuable in developing workable solutions of personnel problems. The approach is primarily utilitarian, with emphasis upon day-to-day personnel problems which confront the aforementioned. The discussions in the book have achieved a high degree of realism through a process of selecting carefully specific problems, techniques and procedures derived from the authors' own experiences and observations. In all the discussions there is no pretense of presenting a detailed manual or a model system of municipal administration.

The tone and outlook of the book in general is objective and practical. The object is not to suggest complete overhauling of personnel administration overnight where obvious disparity exists between accepted standards and practices in given organizations. Instead, the personnel officer is made to realize that changes for the improvement of public service must be introduced by progressive steps.

Progressive municipal officials and personnel psychologists can hardly afford to do without this volume designed as a practical guide. The general administrator will find it very useful in delineating the principal techniques derived by personnel experts and learning the relationship between these techniques and management problems confronting the chief administrator and his department heads. The personnel officer will find the text indispensable for two reasons: first, the principles and activities of a comprehensive personnel program are developed from discussions of the personnel problem; and second, the position of personnel administration in the broader framework of administrative management is portrayed quite effectively. In addition, students of municipal administration, particularly those interested in public service careers, will find the book most profitable in obtaining a balanced picture of the interests and approaches of the administrator and the personnel officer.

John K. McKay

California State Personnel Board

Melville, S. Donald. Color vision, Reprinted from *The Optometric Weekly*, August 24, August 31, September 7, and September 14, 1944, pp. 19.

The author of this article has given a condensed version of the physiology and the psychology of color vision. In most respects the sources cited are up to date and adequately evaluated. However, the field is one in which so much activity is going on currently that one should be careful to accept any crystallization as final. The psychology of color is not given as a psychologist probably would give it, who is particularly interested in the field, but it touches upon most of the facts.

The article works up to a final topic which is probably of considerable interest to readers of the *Weekly*, namely, abnormal color vision and tests for color vision. Most of the current tests are described briefly but no thorough-going evaluation is attempted. The only new material is derived from some tests by the author on the treatment of color blind individuals. His interpretation of these results is in line with other studies which give no indication that color blindness can be altered by dietary or training methods. A list of 89 references is appended, which, while not exhaustive, gives a good sampling of the field. Undoubtedly the article answers very well the purpose for which it was written, namely, that of informing a particular group of professional men concerning a field related to their profession.

Forrest Lee Dimmick

Hobart College,
Geneva, New York

MacKintosh, J. M. *The war and mental health in England.* New York: The Commonwealth Fund, 1944. Pp. 91. $.85.

This book consists of a series of informal essays on mental health in England during the first four years of the present World War and the outlook for the Post-War period. Part One entitled *The impact of war* covers *The process of adjustment, 1939–40, The lonely year, 1940–41, Defense, preparation and alliance, 1941–42,* and *The end of the beginning, 1942–43.* Part Two entitled *Mobilization for peace* covers *Hospital services, Voluntary organizations for mental health, Professional education in mental health,* and *Some problems of the future.* Much of the content consists of either anecdotal material or generalizations for which little supporting data are introduced. The general point of view that scientific services, mental health, education and propaganda can make a great contribution both to war and peace is one with which few will disagree. But the professional person, however much he may enjoy the style and manner of presentation, may well feel that the book is somewhat vague and over-optimistic with regard to the contributions that can be made and quite lacking in concreteness and specificity as to their character.

John E. Anderson

University of Minnesota

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Psychology for the armed services. Edited by Edwin G. Boring. Washington, D. C.: The Infantry Journal, 1945. Pp. 544. $3.00.

Joan chooses occupational therapy. Meta Cobb and Holland Hudson. New York: Dodd, Mead & Co., 1944. Pp. 214. $2.00.

Methods of vocational guidance. Gertrude Forrester. New York 14: D. C. Heath & Co., 1944. Pp. 480. $3.00.

Marriage and family counseling. Sidney E. Goldstein. New York: McGraw-Hill Book Co., 1945. Pp. 457. $3.50.

Dictionary of education. Carter V. Good. New York: McGraw-Hill Book Co., 1945. Pp. 496. $4.00.

Developmental psychology (revised edition). Florence L. Goodenough. New York: D. Appleton-Century Co., 1945. Pp. 723. $3.75.

Job exploration workbook. Milton E. Hahn and Arthur H. Brayfield. Chicago: Science Research Associates. Pp. 95. $.96.

Occupational laboratory manual. Milton E. Hahn and Arthur H. Brayfield. Chicago: Science Research Associates. Pp. 29. $1.00.

Guide to guidance. An annotated bibliography. Volume VII. M. Eunice Hilton. Syracuse: Syracuse University Press, 1945. Pp. 62. $1.00.

Twenty careers of tomorrow. Darrell and Frances Huff. New York: McGraw-Hill Book Co., 1945. Pp. 281. $2.50.

Mainsprings of civilization. Ellsworth Huntington. New York 16: John Wiley & Sons, Inc., 1945. Pp. 660. $4.75.

Mental disorders in later life. Edited by Oscar J. Kaplan. Stanford University: Stanford University Press, 1945. Pp. 436. $5.00.

The governing of men. Alexander H. Leighton. Princeton: Princeton University Press, 1945. Pp. 450. $3.75.

The prediction of success for students in teacher education. Lycia O. Martin. New York: Bureau of Publications, Teachers College, Columbia University, 1945. Pp. 120. $2.00.

Unconsciousness. James G. Miller. New York: John Wiley & Sons, 1942. Pp. 329. $3.00.

Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. Ruth L. Munroe. Stanford University: Stanford University Press, 1945. Pp. 96. $1.25.

Jobs for the physically handicapped. Louise Neuschutz. New York 16: Bernard Ackerman, Inc. Pp. 230. $3.00.

Soldier to civilian. G. K. Pratt. New York: McGraw-Hill Book Co., 1945. Pp. 233. $2.50.

Psychology of sex relations. Theodor Reik. New York: Farrar & Rinehart, Inc., 1945. Pp. 243. $3.00.

Intelligence and its deviations. Mandel Sherman. New York 10: Ronald Press Co., 1945. Pp. 300. $3.75.

Educational psychology revised. Edited by Charles E. Skinner. New York: Prentice-Hall, Inc., 1945. $3.75.

Elementary educational psychology. Edited by Charles E. Skinner. New York: Prentice-Hall, Inc., 1945. $3.25.

Sampling statistics and applications. J. G. Smith and A. J. Duncan. New York: McGraw-Hill Book Co., 1045. Pp. 402. $4.00.

Vocational interest patterns. Irene Wightwick. New York: Bureau of Publications, Teachers College, Columbia University, 1045. (Contributions to Education No. 900.) Pp. 231. $2.60.

An analysis of the work of general clerical employees. New York: Teachers College, Columbia University, 1944. Pp. 100. (Contributions to Education No. 903.)

Employment tests in industry and business. (A bibliography.) Princeton, N. J.: Industrial Relations Section, Princeton University, 1945. Pp. 46. $.50.

Putting the disabled veteran back to work, II. Industrial Hygiene Foundation, Pittsburgh, Pennsylvania. Pp. 33. $.25. (Part I of Proceedings of Ninth Annual Meeting of Industrial Hygiene Foundation of America, Inc., November 15–16, 1944.)

Rehabilitation—a plan to help you employ disabled veterans and other handicapped persons. American Mutual Alliance, 019 N. Michigan Ave., Chicago 11, 1944. Pp. 22. Free.

You and the returning veteran. A guide for foremen. Allis-Chalmers Manufacturing Co., P. O. Box 512, Milwaukee 1, Wisconsin. Pp. 40. Free.

Journal of Applied Psychology

Vol. 29, No. 6 December, 1945

The Accuracy of Precision Instrument Measurement in Industrial Inspection *

C. H. Lawshe, Jr., and Joseph Tiffin

Division of Education and Applied Psychology, Purdue University

Modern industrial production is becoming more and more dependent upon the accuracy of precision instrument inspection. Thousands of employees have been trained in the use of precision measuring instruments, and future industrial developments almost certainly will require still greater emphasis on the accuracy of measurement to insure production which satisfies the fine tolerances of modern precision equipment. Virtually every precision instrument calls upon the operator to exercise judgment in determining proper "feel," "tension," "drag," or other characteristics. In spite of all that is known about the variability of human judgments, little attention has been given to the importance of such variability as it may affect the accuracy of precision instrument measurement. The purpose of the investigation reported in the present paper was to examine the accuracy and variability of employee measurements with certain precision instruments.

Sources of Data. Data were collected in two different plants. The first of these is engaged in the manufacture of variable pitch propellers for aircraft and the second in the manufacture of precision parts for aircraft and automobile engines. There is no evidence that the survey results obtained are any better or any worse than those which would be obtained in other plants of a similar character and there is every reason to believe that similar results would be obtained if the survey were projected to other plants.

An Inspection Department Survey

Job Analysis. Approximately 200 people are employed in the inspection department of the first plant. Their jobs were analyzed by job classifications in order to determine what precision instruments are used and what tolerances are demanded on each job. Frequency counts were then made to determine which instruments or combinations of instruments are used in the largest number of classifications and by the largest number of employees. On the basis of this count, twenty instruments and combinations were chosen as being most important in this particular plant.

---

* The authors acknowledge the assistance of Mr. O. D. Lascoe in establishing the "true" dimensions and of Mr. R. N. Purcell in doing most of the statistical work in connection with the second part of the study.

A Dimensional Control Laboratory. A room was set aside as a dimensional control laboratory and twenty booths or inspection stations were set up. Each booth was numbered and in it were placed one of the twenty instruments, a standard part from the plant, and a simplified working drawing which indicated one dimension to be measured with the instrument provided. When an employee entered the room, the attendant determined his job classification and provided him with an appropriate work-sheet for each of the stations containing work samples from his job. Each employee was tested on only those instruments which he uses on his particular job. He was encouraged to make five measurements and then to record his best judgment as to the dimension. The readings thus obtained were compared with so-called "true" dimensions which were determined by means of ultra-precision instruments in combination with Johansen blocks. Instruments utilized in the performance testing were checked and adjusted periodically to insure constancy.

Emphasis of Testing. This performance testing procedure was organized in connection with a training program and its primary function was to identify persons in need of training. Plans for a maintenance program were also made with provisions for re-testing employees every three months. It was also planned to utilize the laboratory to supplement seniority in determining adequacy in connection with transfers and promotions. The program was instituted with the knowledge and backing of line supervision and of the union in the plant. There is every reason to believe that nearly all of the employees approaches the test situation with a favorable attitude.

Results

Results obtained at eleven of the twenty stations are presented in Figure 1. The particular stations selected for illustration were chosen in terms of general familiarity with the instruments used and not because of any peculiarity in the findings; they are truly representative.

In Figure 1, the open bars represent the percentage of inspectors tested who obtained readings within the established tolerance. The solid bars represent the percentage of persons tested who failed to meet the standard. As already stated, not all of the inspectors were tested at each station; instead, the sample contains only those who use the instruments on their jobs. This accounts for the fact that the N's range from 117 to 162. The figure indicates that the percentage of inspectors meeting the various standards ranged from a high of 66% on the inside microm-

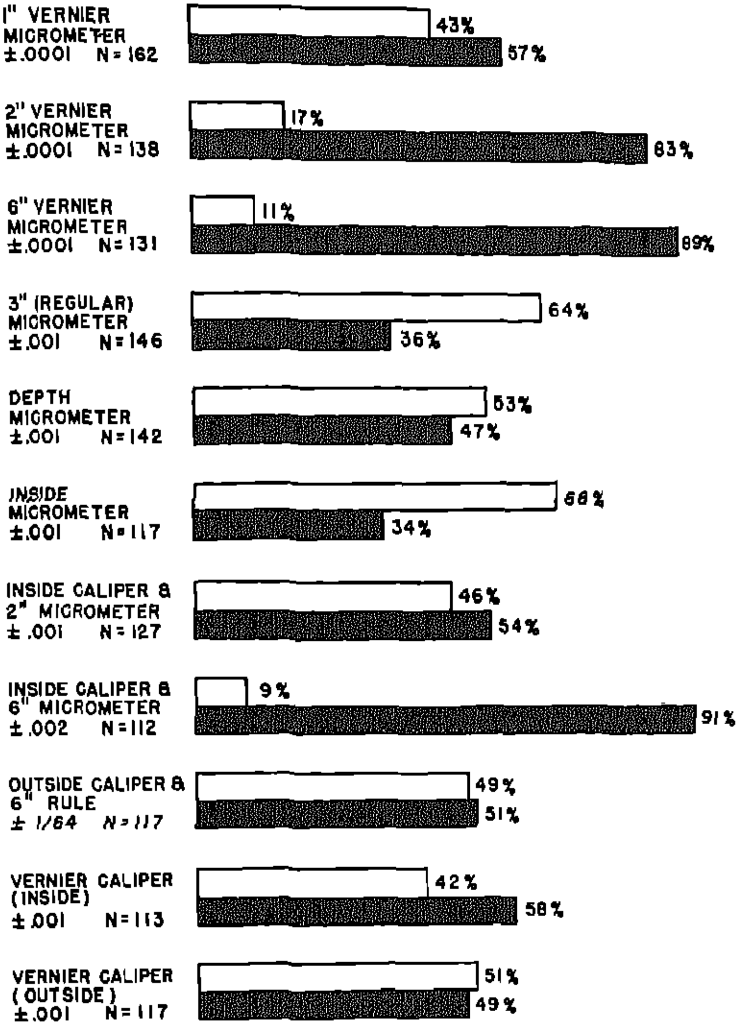


FIG. 1. The percentage of inspectors passing and failing various precision measuring instrument performance tests in an aircraft propeller plant. The open bars indicate the percentage meeting the standard and the solid bars indicate the percentage failing.

eter to a low of 9% on the inside caliper in combination with the six-inch micrometer. The pattern of performance on the various vernier micrometers also seems significant. It will be noted that 43% of those tested met the standard with the one-inch micrometer, 17% with the two-inch, and only 11% with the six-inch. The varying tolerances established for the instruments are the same as the tolerances which job analyses indicated had been established by the engineering department and are identical with those encountered in the shop.

A Tool Room Survey

Procedure. Because many of the employees in the plant just described were new and were drawn from a "tight" labor market, a related study dealing with experienced toolmakers was set up in a plant engaged in the manufacture of precision parts for aircraft and automobile engines. In this study, 45 men were selected from the tool room. Their ages ranged from 17 to 62 years, their experience with the company from five to twenty-nine months, and their experience on their present jobs from one to twenty-nine months. For the most part, the job classifications of these men fall in higher labor grades than do those reported in the inspection department study.

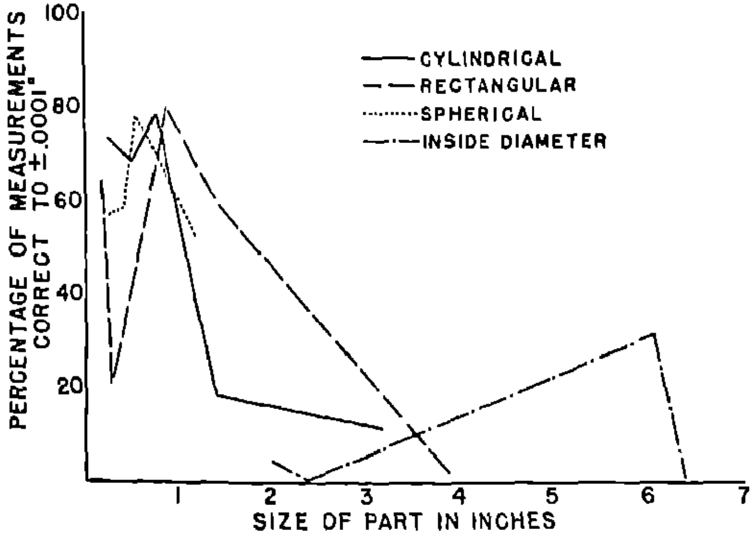


FIG. 2. The percentage of toolmakers ($N = 45$) who obtained vernier micrometer measurements within .0001 inch of the "true" dimension on each of nineteen parts. The base line indicates the approximate size of the dimension and the code indicates the shape of the part measured.

The study was limited to the use of vernier micrometers and employed nineteen parts to be measured. Five parts were cylindrical, five rectangular, five spherical, and four were inside diameters. Here again, each employee measured each part independently five times and the reading recorded was the best judgment he could make as to the "true" reading on the basis of these trials. After all readings had been completed by all of the men, the parts were measured with ultra-precision instruments and Johansen blocks in order to obtain the closest possible approximations to the "true" dimensions against which to compare the measurements made by the men.

Accuracy Results. No significant correlations were found between accuracy of measurement and either age, length of experience with the company, or amount of time on the present job. The percentages of accuracy are shown in Figure 2. On the baseline of this figure is plotted the approximate size of the part. On the vertical axis is plotted the percentage of readings correct to .0001 inch. It will be noted that the parts vary in size from approximately ¼ inch to approximately 6½ inches. Each of the four lines plotted in Figure 2 shows the percentage of readings within .0001 inch of the "true" dimensions for parts of a certain shape according to the code indicated in the figure. Thus, for cylindrical parts, about ¼ inch in size, 73% of the readings were accurate to .0001 inch. For cylindrical parts approximately 3 inches in size, however, only 12% of the readings were accurate to .0001 inch.

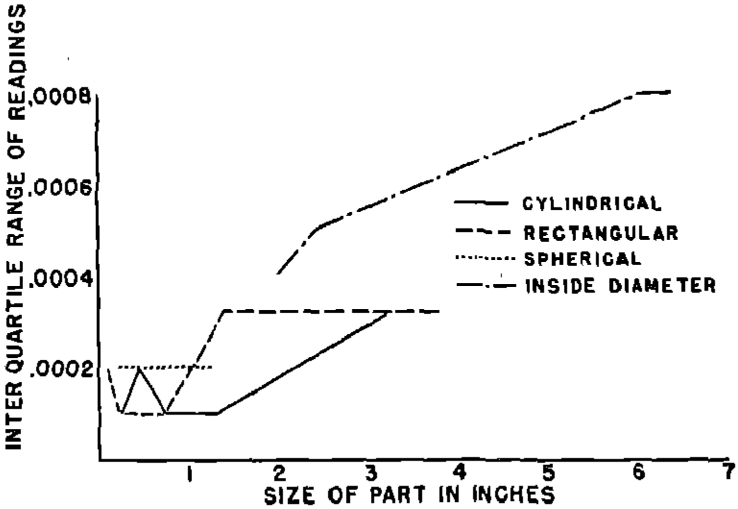


FIG. 3. The variability of the vernier micrometer measurements of 45 toolmakers on each of nineteen parts. The base line indicates the approximate size of the dimension and the code indicates the shape of the part measured.

Variability Results. The results plotted in Figure 2 assume the accuracy of the so-called "true" dimension. The validity of these "true" dimensions is always open to question in spite of the ultra-precision methods used. Therefore the data were analyzed in another way to show the variability of the readings obtained without reference to the "true" dimensions. This analysis of the results is plotted in Figure 3. Here again the approximate size of the part in inches is plotted on the baseline. The vertical axis on this chart plots the interquartile range of the readings. This may be interpreted to mean the range of the middle 50% of the readings under each condition of size and shape of the part. For example, looking at the solid line plotted in Figure 3, it will be noted that when the part is cylindrical and approximately ½ inch in size the interquartile range of readings, or the middle 50%, is .0002 inch. This means that 50% of the readings were within a range of .0002 whereas the remaining 50% of the readings varied by more than .0002, either above or below of the range of the middle 50%. In like fashion, it will be noted that when the part is cylindrical and approximately 2¼ inches in size, the middle 50% of the readings fall within a range of .0003, whereas the remaining 50% of the readings on this part fall more than .0003, one way or the other, from this range.

Summary and Conclusions

Controlled performance tests with various precision instruments were administered in two industrial plants. In one plant 200 inspectors were tested on a variety of instruments used in their respective jobs. In the other, 45 tool room employees were tested on vernier micrometers with parts of varying sizes and shapes.

In general, the following conclusions are supported:

1. The accuracy of precision instrument usage is probably considerably lower than is ordinarily assumed by those responsible for methods and standards. In one plant the percentage of inspectors meeting the standard ranged from 66% on the use of the inside micrometer to 9% on the inside caliper in combination with the six-inch micrometer.

2. In the population studied, micrometer reading accuracy did not correlate significantly with age, amount of experience with the company, or length of time on the present job.

3. Gross size of the part is apparently a factor in the accuracy of micrometer measurement. Under optimal conditions in the second plant, not more than 80% of the readings were accurate to .0001 inch and with larger parts, ranging from 3 to 6 inches, only about 20% of the readings were accurate within these limits.

4. Gross size of the dimension is likewise related to the variability of measurements. As the parts increase in size, regardless of shape, the spread of the readings becomes greater so that for large inside diameters, 50% of the readings vary by .0008 inch from the other 50% of the readings.

5. The results suggest that while inside diameter measurements are more variable than measurements of cylindrical, rectangular, or spherical dimensions, the percentage of measurements meeting the standard of ±.0001 inch is no less. However, the problem of the relationship between shape and both accuracy and variability is open to further investigation.

6. The necessity for the development of training methods that will more nearly standardize judgments based on such characteristics as "feel," "tension," and "drag" in the use of precision instruments is indicated.

7. The implication is present that the very nature of the vernier micrometer and similar precision measuring instruments is such that one should not expect as high a degree of constancy as the average operator, supervisor, and standards man has been taught to expect.

Received November 29, 1944.

Movement Analysis as an Industrial Training Method *

Lawrence G. Lindahl

Division of Education and Applied Psychology, Purdue University

Many industrial jobs which involve coordination of hand and foot movements in operating machines present difficult training problems. Typical of such jobs is contact disc cutting. The company producing contact discs experienced considerable difficulty in training new operators because the majority did not complete the training and for those who did, learning was slow and uncertain. Preliminary investigation with experimental apparatus indicated that part of the difficulty encountered in training new operative personnel was due to the failure to identify the form of cutting movement necessary for "getting the feel" of satisfactory performance of the job.

The cutoff machine slices thin discs (e.g., .020" × .150" diameter with ±.002" on both dimensions) from various sizes of tungsten rods with a rubber-bonded abrasive wheel .015 of an inch thick and six inches in diameter. The cutting wheel turns between closely fitted guides, moving up and down between the guides and across the rods being cut. The process is wet cutting and the wheel is not visible to the operator while cutting. Most machines cut two rods at a time.

The operator pushes the rods through guides into stops which regulate the thickness of the discs. Holding the rods firmly against the stops with a hand lever, he operates a pedal with the left foot which controls the cutoff wheel as it is applied to the rods. As soon as he cuts through the rods he lifts his foot, the wheel rises, and immediately after it has cleared the stops a backward jerk of the right hand actuates ejector or knock-out pins which knock the severed discs into the stream of water. Immediately after the ejection of the discs the rods are pushed back into the stops and a new cut is taken. Figure 1 shows an operator seated at a machine ready for operating.

The cutoff machine depends for successful operation upon the speed, form, rhythm, and pressure pattern of the hand and foot action of the operator. Failure to apply foot pressure properly results in damage to the discs, excessive breakage and use of wheels, and wastage of material.

---

* Based upon a thesis submitted by Lawrence Gaylord Lindahl to the Faculty of Purdue University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, October 1944. Acknowledgment is due Dr. Joseph Tiffin and Dr. C. H. Lawshe who jointly directed the research and to James R. Brock who made the study possible in industry.

The purpose of this study was to analyze the disc cutting operation by identifying the form of the foot movement that produced satisfactory quantity and quality of discs with minimum cutting wheel usage, and to teach this form to new operators by the movement analysis method.

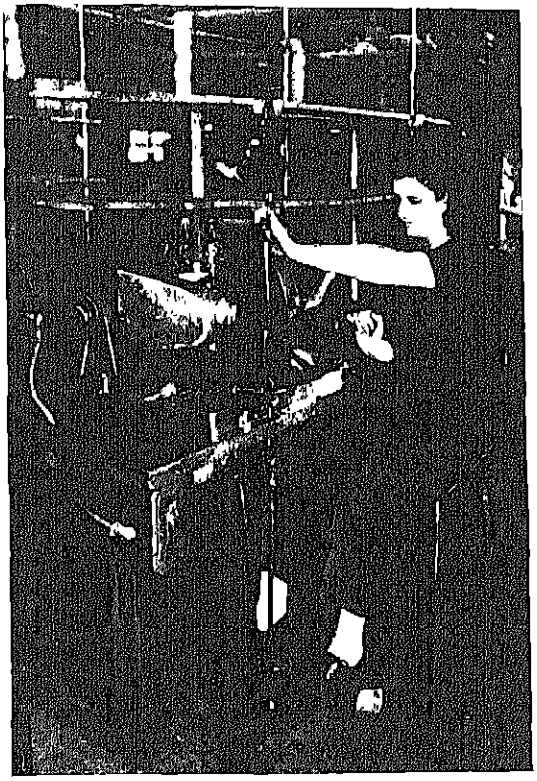


FIG. 1. An operator seated at the disc cutoff machine ready for operating.

The Method

The Apparatus. The mechanical apparatus used for the movement analysis included a paper tape recorder with a specially made writing arm attached. A fine thread was fastened to the writing arm and run through glass bushings mounted in metal pieces which were clamped at accessible places on the cutoff machines, and thence to the pedal in such a manner that the complete cutting cycle movement could be recorded on the paper tape as it moved along at a known speed under the pen.

The recorder was placed on a small wheeled table which was pushed from machine to machine and checks on operators were quickly and accurately made without interference with production. Figure 2 is a schematic drawing of the recorder attached to the cutoff machine.

Procedure. The procedure consisted first of a job analysis by activity. The principal activity, the cutting cycle, was further analyzed by studying the foot action recordings of skilled operators to find out what constituted good performance. The recorder was then used for instructional purposes in the training of new operators and for improving experienced operators already on the line. This type of analysis is not entirely new. It was utilized with good results by Tiffin and Rogers (5) in training tin plate inspectors and by English (1) in training riflemen.

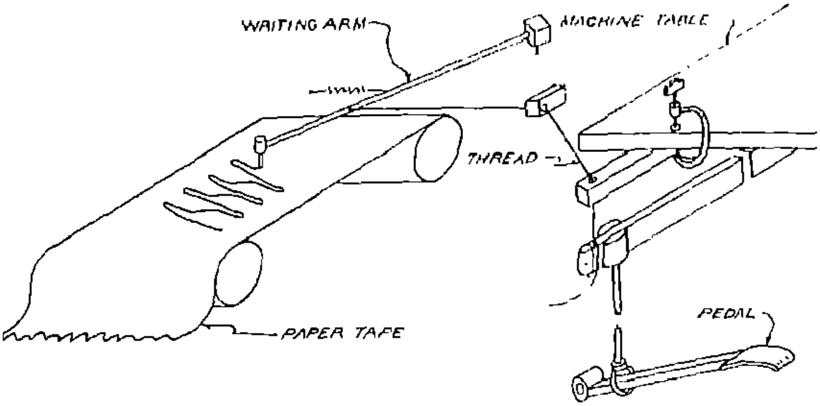


FIG. 2. Schematic drawing of the paper tape recorder attached to disc cutoff machine

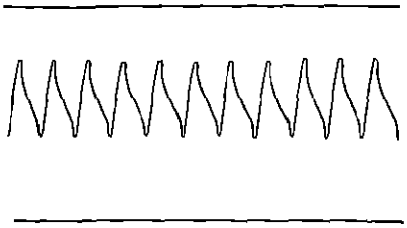


FIG. 3. Disc cutter foot-action pattern of a good, experienced operator. This was the accepted standard.

The coordination of the foot action in cutting, with the hand action in placing the rods against the stops and ejecting the discs after cutting, constituted the cutting cycle. The cut through the rod was the cutting phase and the ejecting of the discs and placing of the rods against the stops was the recovery phase.

Since the foot action was the principal part of the cutting cycle, the main effort was given to finding the correct foot action pattern and then presenting it to the trainees in an effective manner.

By recording the patterns of experienced operators and supervisors, who were also experienced cutters, and comparing the action patterns with quantity, quality records, and abrasive wheel usage records, it was possible to identify the "standard" or correct pattern. There were wide variations in action patterns of the experienced operators showing the existence of individual differences, but the consistent pattern was there to form the basis on which the noted differences could be explained. Figure 3 shows the foot action pattern accepted as standard. All sizes of rods gave the same pattern, but spacing of cuts was wider for large rods because of longer cutting time. The correct pattern showed a good reverse cutting curve which indicated unhurried, steady cutting with "feel" of rod. The operator checked his foot at the same place on every cut just as he cut through the rod by easing the pressure, thus allowing the wheel to cut its way through without making burrs. The short follow-through at the bottom indicated, in shop parlance, a "soft" foot. The recovery started soon after cutting through, saving time for each cut

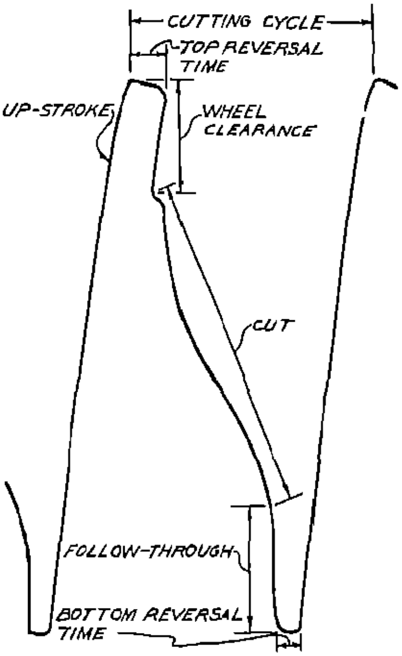


Fig. 4. An enlarged view of disc cutting cycle showing its principal parts.

and allowing plenty of time to coordinate or time the ejection of the discs and place the rods back against the stops ready for the next cut. The operator paced the machine and adjusted his time for the cutting part of the cycle to the speed at which the wheel could cut best without forcing or crowding the wheel. To force or crowd the wheel results in several disc defects, short wheel life, and wear out of the setup. Figure 4 is an enlarged view of one cycle showing the significance of each part of the cycle. Figure 5 is an incorrect pattern made by a trainee with only 4 hours experience. Figure 6 shows the foot-action patterns of one trainee at various stages in the training program. There is a steady approach to the standard pattern and a complete story is evident in each recording after the various hours of supervised operation. It will be noted that the pattern at 239 hours closely resembles the standard pattern shown in Figure 3.

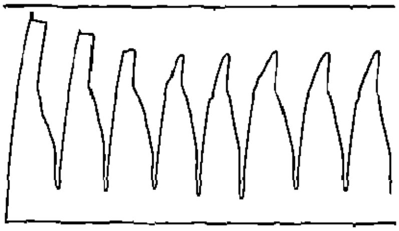


FIG. 5. An incorrect disc cutter foot-action pattern. Operator had only 4 hours experience.

Enlarged instructional posters with analytical notations were prepared both for "correct" and "incorrect" patterns and for various types of damage to the product which were shown to be reflected in the action pattern. Individual action patterns of each operator were kept in folder form also so the operators could note their progress. By careful use of the recorder at timely intervals with both the trainees and experienced operators and by interpreting the foot action patterns in terms of the standard, it was possible to reduce the training time of new operators and to improve the quality and quantity performance of some employees already on the job.

Results

Trainees Versus Old Operators

Training a group of operators on the production line is not unlike breaking in a team of horses, a group of men for a football team, or a squad for the army. All must work together. Especially is this true where the operators are working on day rate and a certain amount of production must be obtained each day. One individual cannot do all the work no matter how good he is. The object of the training program is, *therefore, to get all of the trainees* as nearly alike as possible in the shortest time and all as good and as near like the standard as possible.

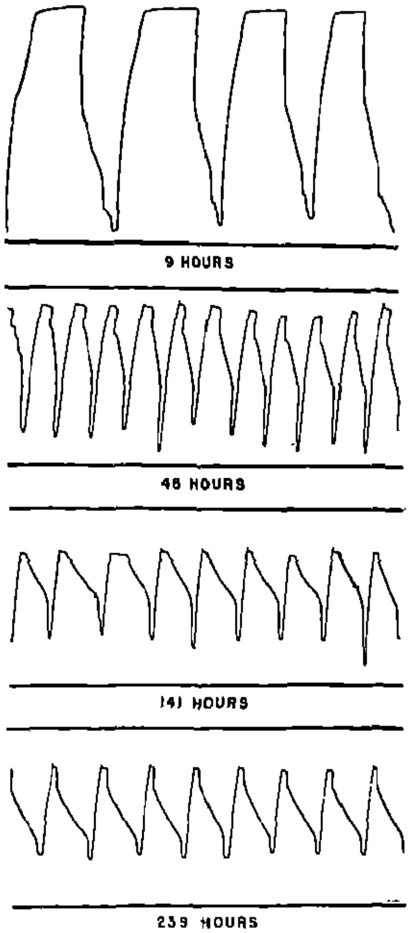


Fig. 6. Disc cutter foot-action patterns of a trainee showing improvement with training. The records were made after 9, 45, 141, and 239 hours of supervised operation.

The supervisor, or trainer, has to move from operator to operator, checking, helping, comparing, encouraging, correcting, explaining, and expecting progress day by day. If he does not maintain his own interest he will not find continued interest in the learners, especially after the newness of the job has worn off. He should be *interested in the group;*

it is from the group that he expects to obtain production. It is, therefore, on the group that the training program must be evaluated. Even when all possible attention has been given to the group, individual differences will remain. Tiffin (6) enumerates many kinds of individual differences in job qualifications and individual productivity. The person doing the training must constantly detect and give due consideration to these individual differences in order to get the entire group to the stage of perfection he desires. Equal increments of practice and training do not in any sense reduce all trainees to a common level of performance. In evaluating the results of a training method on the job where none of the variables is controlled, there are factors that prevent one from giving too much attention to the production of any one individual. For instance, there are external limitations on the individual's power to increase his output, such as the condition and speed of his machine. One trainee may be given credit for less output than another because he was carrying an undetected handicap. And even though this handicap is detected there is often no way of eliminating it. The speed of two machines that look exactly alike may vary tremendously during different periods of the day; or supplies of material or the difficulty of the task may set an upper limit on individual production which perhaps one-third of the members of the group could reach, and no one could better, no matter how competent he might be. Age is another factor which has unmeasured effects on individual production on the line and makes it necessary to consider group differences in evaluating the results of the training program.

There are numerous ways to evaluate the results of the training once it has been put into operation. Lawshe (2) lists 13 methods and from such an extensive list one should find a method that fits any particular case.

In machine operations, like disc cutting, jobs have been time-studied, rates based on these studies have been set, and production of trainees and old workers alike is indicated by the output. In the particular operation reported herein training can be evaluated on two factors, namely, production performance and wheel performance. Both wheel breakage and wheel use are important aspects of wheel performance. The two factors are considered to be of equal importance and management has set up a rate sheet based on the number of cuts secured per wheel for a given size of rod and the speed of cutting (production per hour). In other words, the operator must get a specified number of cuts per wheel and must cut at such a speed that he will turn out enough units in one hour to earn the base rate set per hour. When he gets the exact number of cuts per wheel for the rod size he is cutting and cuts just enough to earn his hourly rate, he is doing 100% performance in both factors, namely, production performance and wheel performance.

There are several reasons for emphasizing wheel performance. One is that the special type of wheel used for tungsten disc cutting is very expensive and wheel costs can very easily exceed labor costs. Also, a better quality of product is obtained by limiting the speed of cutting, which is effectively done by placing emphasis on wheel performance.

Curves are used to evaluate the effects of this training program. As stated by Tiffin (6, 186), "such curves, therefore, serve the very useful purpose of providing a means of evaluating the successfulness of an operator-training program and spotting decisively those operations in which training is inadequate, either in quantity or quality."

Figure 7 is the production curve of the trainees for 12 weeks. Production performance percentage is plotted on the vertical axis and weeks (6 days each) of training on the horizontal axis. Production performance percentage makes the production performance of all operators comparable since it is computed on the company rate sheet for the various sized rods cut. For instance, if an operator using 6″ wheels to cut .142 inch rods, cuts 12,000 units of rod in 10 hours, he cuts at the rate of 1,200 units per hour. The rate sheet indicates the operator should cut 2,340 units per hour to earn the base rate of 65 cents per hour. Since 1,200 is 51 per cent of 2,340, his production performance is 51 per cent. The production performance percentage is, therefore, the average number of units cut per hour divided by the number of units that the operator is expected to cut per hour of the size rod being cut to earn his wage.

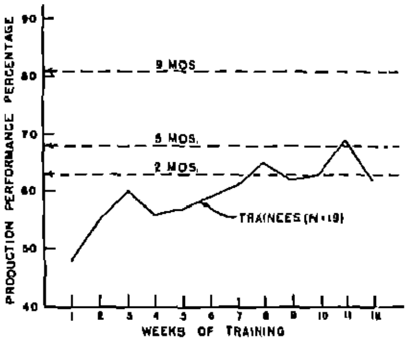


FIG. 7. Production performance percentage of disc cutter trainees. Dotted lines represent average production performance percentage for first 2 weeks of groups with an average of 2, 5, and 9 months experience at start of training.

The dotted lines in Figure 7 show the production performance of old operators at the start of the training experiment. There are three groups of these old operators. Before the training started, eight operators had an average of 2 months, nine operators had an average of 5 months, and ten operators 9 months experience on the job. The trainees had no previous experience. Old operators total 27 and trainees 19. No data were available on the production abilities of the various groups prior to the start of training. In order to obtain a basis of comparison for the groups, an average of the production of the first two weeks at the beginning of the training is considered indicative of the ability of the groups of old operators at that time. These averages are shown on the figure by the dotted lines.

The production curve for the trainees reflects a steady rise. At the end of 8 weeks they have surpassed the average of the 2 months group. At 11 weeks they exceeded what the 5 months group was able to do at the beginning of the training program. Although their average fell on the 12th week, this is not unusual in curves of this type. The general trend is upward and with more confidence, experience, and continued training it should be safe to assume that the curve should reach the top limit of the 9 months group in much less time than 9 months, resulting in saving many man hours of training time.

One may legitimately inquire about the quality of the discs cut. There was no difference in the quality of the discs cut by the trainees and by the other groups. The company used a production checker who continually checked the discs as they were cut by going from operator to operator and it was practically impossible for any one operator to cut a majority of the defective discs. The machine would be shut down or the operator's performance investigated to determine the cause of the defects. Any shut down would be reflected in the production performance of the particular operator. Since tungsten is an expensive material, it was imperative to keep wastage at the very minimum. One hundred per cent inspection of the discs was made after a tumbling operation, but this was done according to size of discs and the production of two or three operators cutting the same size rods was usually thrown together. Production control regulated the quantity to be cut over any one period. In the form of stricter supervision this control probably served as a forced incentive to the worker.

Figure 8 is the wheel performance percentage plotted against weeks of training. The averages of the 2, 5, and 9 months groups of old operators in wheel performance are shown by the dotted lines. These averages are of the first two weeks performance at the beginning of the program. Wheel performance percentage is computed from the rate sheet. Referring to the illustration previously cited, if the operator who cut 12,000 discs in 10 hours used, without breaking, 20 of the 6″ wheels and broke 5 others, he used a total of 25 wheels to cut 12,000 discs. He obtained 480 units per wheel. According to the rate sheet, 100 per cent wheel performance on .142 rod requires 470 units per wheel. The wheel performance is 480 divided by 470 or 104 per cent. Wheel performance is, therefore, the units the operator gets from each wheel divided by the number he should obtain according to the rate sheet for the size of rod being cut. Wheels broken must be counted as wheels used, for the operator gets some cuts with the wheels before they are broken. Wheel performance is negatively correlated with production performance. The faster the cutting the higher the production and the fewer units per wheel. This relationship had to be regulated in the training program and emphasis placed on cutting according to the cutting curve so that both maximum wheel performance and maximum production were obtained.

The rate sheet is made up on the basis of 100 per cent production performance and 100 per cent wheel performance. Figure 8 shows the 9 months operators at exactly 100 per cent wheel performance while Figure 7 shows their production performance at 81 per cent. It will be observed by reference to Figures 7 and 8 that the trainee curve goes up toward 100 per cent production performance and downward toward 100 per cent wheel performance.

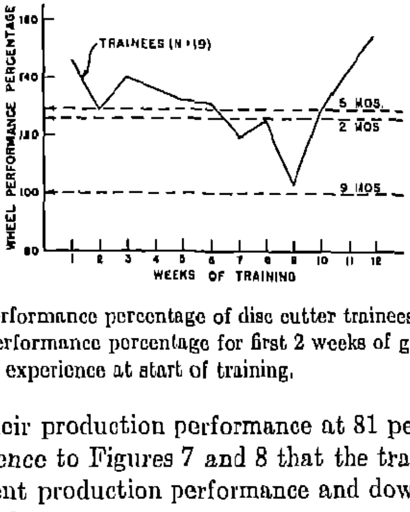


FIG. 8. Wheel performance percentage of disc cutter trainees. Dotted lines represent average wheel performance percentage for first 2 weeks of groups with an average of 2, 5, and 9 months experience at start of training.

One must speculate on the sudden upturn in the wheel performance curve of the trainees after the 9th week. It can be accounted for to some extent at least by the fact that a more rigid accounting of wheels used was started at that time. Some operators had acquired the habit of not using all the wheel because of the slower cutting pace due to the smaller diameter of the wheel as it wore down. These operators changed wheels before entirely using them. This lowered the wheel performance and took more wheels. By requiring operators to turn in every used wheel for inspection at the end of the day's run it was possible to determine just which operators had not entirely used their wheels. The general effect was that all operators tended to use up their wheels, thus getting higher wheel performance. This might also cause the slightly lower production because of slower cutting as the wheel grew smaller.

The best operator would get as much over 100 per cent wheel performance and as much over 100 per cent production performance as possible. To do this he would have to cut steadily, consistently, and carefully during all working time. Failing to do this he could not possibly get over 100 per cent wheel performance even though he might cut fast to make up lost production time, because the faster he would cut the lower his wheel performance would become. The interrelationship of wheel performance to production performance complicated the training problem and made the movement analysis method all the more important.

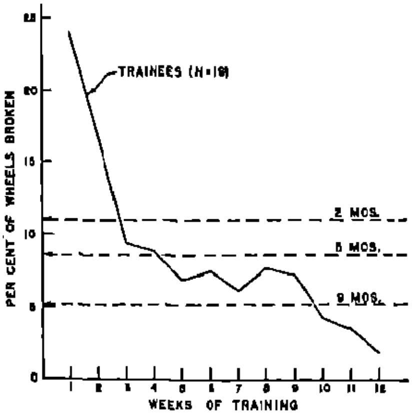


Fig. 9. Percentage of wheels broken by disc cutter trainees. Dotted lines represent average percentage of wheels broken for first 2 weeks by groups with an average of 2, 5, and 9 months experience at start of training.

As has been noted, this found a standard pattern that laid stress on both accuracy and wheel performance.

Figure 9 shows the percentage of wheels broken plotted against weeks of time. Percentage of wheels broken is the fraction of the total number of wheels used each week. For instance, the number of wheels used without breaking plus the number broken is the total number used. The ratio of the number broken to this total is the percentage broken. If, for example, an operator wore out 300 wheels in one week and broke an additional 25, he used a total of 325 wheels. The percentage of wheels broken during the week is 25 divided by 325, or 7.7.

Figure 9 shows graphically the trainees' reduction in wheel breakage due to the attention given to the correct cutting method. During the first week the trainees broke 24 per cent of their wheels. By the third week they were breaking less than operators having 2 months experience, by the 5th week less than those having 5 months experience, and by the 10th week less than those having 9 months experience. By the 12th week of training the trainees broke only 1.8 per cent of their wheels, or about one-third that of the group with 9 months previous experience. This is good for two reasons; first, because the wheels were very expensive and large savings resulted (wheel cost could very easily exceed labor cost), and second, all time taken to change broken wheels has to be charged against production time. Broken wheels prevent running for a period ranging from one minute to sometimes as long as a half hour, depending on how long it takes to remove the broken pieces from between the guides.

The higher average wheel performance noted for the trainees is due to emphasis being placed on accuracy rather than speed and on correct operation, or cutting according to the reverse cutting curve. High wheel performance eventually leads to increased production. It insures a better quality of product at the beginning with less wastage. The curve shows that the trainees steadily reduced their wheel breakage. In Figures 7, 8, and 9 the trainees show exceptionally fine performance in relationship to the other three groups of operators who learned by the "pick-up" method. Although their production average was lower than the 9 months group of operators, they steadily increased production, their wheel performance was maintained well over 100 per cent, and the per cent of wheels broken was considerably reduced.

Trainees Versus Beginners

Another way operators learn jobs, probably more prevalent than any other learning method in industry, is that of "trial and success." This consists of showing the operator his job and then "turning him loose" to learn in the best way he can. It is a sort of "sink or swim" proposition. The operator is not led from the beginning by progressive steps and has no reliable indication of his progress. Such haphazard learning usually results in lower efficiency and poor work habits.

It was possible to obtain records over a period of three weeks of the production and the wheel performance of 5 operators who were learning the disc cutting operation in just this manner. They are called beginners for purposes of comparison with the trainees who were trained by an organized method. These beginners did not have the advantage of using the movement analysis recorder nor any of the instructional material.

Figures 10, 11, and 12 present a comparison of the production performance, wheel performance, and wheel breakage of beginners and trainees at the end of the first, second, and third week of operation on the job. Figures 10 and 12 do not show a significant difference between the two groups in production performance and percentage of wheels broken. Figure 11, however, shows graphically considerable difference in wheel performance percentage between the two groups. Whether the difference shown can be attributed to chance can be tested statistically by finding (*t*) the significance ratio for the difference in the means of two independent small random samples. The smallest difference between the two curves occurred at the second week. The significance of this difference, therefore, was computed to determine the possible role of chance in accounting for the difference between the curves. The second week of training or experience shows a wheel performance percentage mean of 129 for the trainees and 86 for the beginners. Substitution in the formula (3) for finding the significance ratio for the difference in the means of two independent small random samples results in a "*t*" of 2.548.

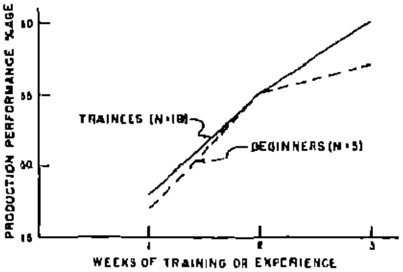


FIG. 10. Production performance percentage of disc cutter trainees and beginners for first three weeks.

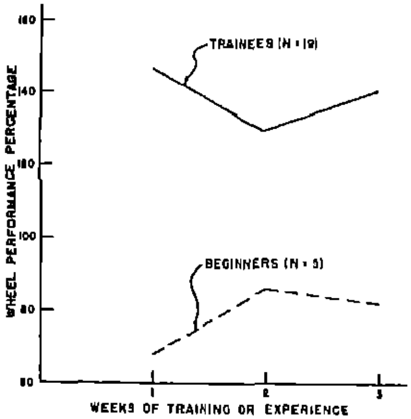


FIG. 11. Wheel performance percentage of disc cutter trainees and beginners for first three weeks.

Reference to the table of *t*'s indicates a "*t*" of 2.518 for 21 degrees of freedom is required for the 2 per cent confidence level. The obtained difference is therefore significant at the 2 per cent level which means that there are 98 chances in 100 that the difference found is not due to chance and must therefore be due to the special training given the experimental group.

Effect of Training Program on Old Operators

When any training is carried out directly on the production line it is bound to have some effect on the experienced operators even though most of the attention is directed to the trainees. Studies have been made of factory workers in which work conditions of all kinds were varied. A noteworthy example is the experiment carried out in the Hawthorne plant of the Western Electric Company (4). Its most important finding was that regardless of the nature of the changes made in the working conditions the productivity of the experimental group tended in general to increase. Although the interest and attention undoubtedly had a stimulating effect on the old operators, the fact that foot-action recordings and interpretations were made for them as well as the trainees would lead one to believe that interest and attention did not account for the total improvement.

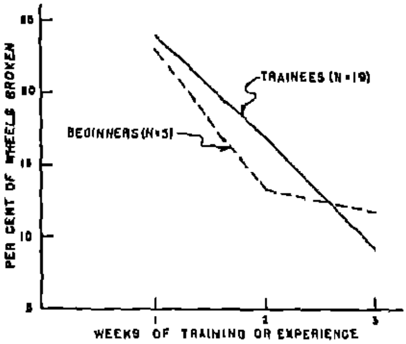


FIG. 12. Percentage of wheels broken by disc cutter trainees and beginners for first three weeks.

One cannot rush up to an old operator with a new gadget with the expressed purpose of checking up on him. The cooperation of the old operator may be secured, however, by explaining to him the need for finding out how the job is done so that it can be taught to the new operators. Practically every old operator will give the best performance he can muster up and is then already on the road to self improvement. One can also question the operator about his work. As soon as the old operator finds that someone is interested in his job he begins to think that the job is important after all and he is likely to take more interest in his work and that of fellow workers. Learning becomes contagious, production starts to increase, the entire line improves, and no one but the person in charge of the training is aware of it or knows the reason why.

If those operators who seem not to want help are ignored for a while they will soon seek help either by asking for a recording or by letting it be known that their recordings have not been taken. Workers all desire to be treated alike.

A review of the foot-action patterns[1] shows that old operators used a lot of waste motion in cutting and most of them had not developed the necessary "feel." That they could improve by watching the recorder was also shown.

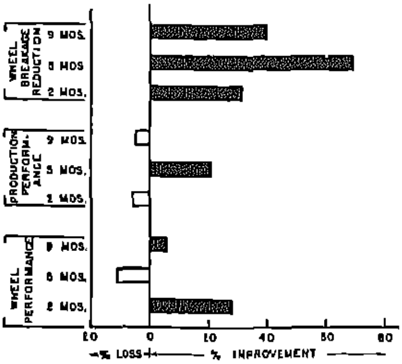


FIG. 13. Percentage of improvement or loss in wheel performance, production performance, and wheel breakage reduction for groups having an average of 2, 5, and 9 months experience on the disc cutoff machine prior to the start of the training program.

Figure 13 shows the percentage of improvement or loss in wheel performance, production performance, and wheel breakage reduction made during training by groups having an average of 2, 5, and 9 months of experience on the disc cutoff machine prior to the start of the program. The graph is based on averages for the first three weeks and the last three weeks of the training period. In computing the percentage of improvement or loss, the performance at the start of the program was considered 100 per cent.

All three of the groups broke fewer wheels after the training was under way. The 5 months group made the greatest improvement by breaking 69 per cent fewer wheels than they did at the start, the 9 months group

---

[1] Case studies and complete presentation of data will be found in the appendix of a thesis by the author on file in the library of Purdue University.

40 per cent less, and the 2 months group 32 per cent less. Reduction in wheel breakage effected a substantial saving in wheel costs over the period.

The 5 months group improved their production performance 21 per cent but suffered a loss of 11 per cent in wheel performance. Since production performance is negatively correlated with wheel performance and both are of equal importance, this represents total improvement and reflects the emphasis placed on wheel performance by the movement analysis method. Likewise the 2 and 9 months groups improved their wheel performance by 28 and 6 per cent, respectively, at the expense of production by only 6 and 5 per cent. Again with wheel performance and production performance of equal importance this represents substantial improvement.

The total picture is good when one considers the nature of the operation, that it is a hand and foot coordinated operation which requires the human operator to function with mechanical precision over long periods of time, and that old operators may have learned as many bad habits as good, necessitating unlearning before starting new habits.

That these experienced operators could so effectively cut down in the number of wheels broken and improve in both wheel and production performance at this stage of experience indicates that learning of a better cutting cycle by old, experienced operators took place with the installation of the program for training new operators. The improvement made by the experienced operators as well as the trainees seems to justify movement analysis as an effective industrial training method.

Summary and Conclusions

The problem of training rapidly a number of new workers to operate certain type of cutoff machines used for cutting tungsten rods into small discs for electrical apparatus has been explained. A method has been described which analyzes the foot movement of the operator by the recording of a pattern on a moving paper tape. A standard pattern was established and the recorder was used to take recordings of the trainees at frequent intervals. These recordings were explained on the basis of the standard. Various defects in the product were identified with the foot movement patterns and many individual differences in both old and new operators were noted and described as clinical case studies. Finally, complete production records of all operators were kept. These records were plotted in graphical form to show the progress of the various groups at different stages of training.

In general, the findings warrant the following conclusions:

1. The disc cutting operation was analyzed by activity and the best form of cutting movement identified.

2. The cutting movement of the operators was shown in graphic form, compared with a standard, and used as an effective training method.

3. Training time was effectively reduced. At 8 weeks the production performance percentage of the trainees was better than that of old operators who had had the same average amount of experience by the "pick-up" method.

4. Trainees, who were taught to use the cutting wheel by studying the movement analysis recordings, secured better wheel performance than old operators.

5. Trainees reduced their wheel breakage in a few weeks to a point much lower than the average of old operators at start of training program.

6. Trainees did better in wheel performance than beginners who learned by the "pick-up" method.

7. Experienced operators benefited by the training program.

8. It is entirely possible that equally successful improvements in training could be made in numerous other manipulative jobs by the application of a similar method of recording graphically the performance of the operator.

Received November 1, 1944.

References

1. English, H. G. How psychology can facilitate military training. *J. appl. Psychol.*, 1942, 26, 3-7.
2. Lawshe, C. H., Jr. Training operative personnel. *J. consult. Psychol.*, 1944, 8, 158.
3. Lindquist, E. F. *A first course in statistics.* New York: Houghton Mifflin Company, 1942, pp. 138-139, 240.
4. Mayo, E. *The human problems of an industrial civilization.* New York: The Macmillan Company, 1933, Chs. 3-5.
5. Tiffin, J., and Rogers, H. B. The selection and training of inspectors. *Personnel*, 1943, 22, 3-20.
6. Tiffin, J. *Industrial psychology.* New York; Prentice-Hall, Inc., 1942, pp. 1-10.

Military Job Evaluation: Army Air Forces *

Arthur O. England,** Captain, Air Corps

During the early stages of 1942 and 1943, when the reception centers were bulging with incoming men, the paramount problems facing the classification and personnel officers were the initial classifying of these men, fulfilling the technical school quotas and inaugurating on-the-job training necessary to build efficiently operating tactical units. Little or no information was available on the subject of effective utilization of manpower. Job evaluation, by any method, was non-existent—or if existent, at least it was not available for general consumption by military personnel.

A War Department Memorandum [1] issued in January 1943 stated in part that many enlisted men in the Army Air Forces who have been trained as technicians in the various Army Air Forces Technical Schools, were being employed on jobs which did not require special training or special qualifications and which were appropriate for unskilled privates only. The Inspector General's Department assumed some of the responsibility for detecting and reporting to the War Department personnel misplacement and attendant wastage. Further, every Air Force or Command established a special unit known as Classification Audit Teams, whose function it is to inspect periodically each installation in that Command for proper classification, assignment and personnel utilization. These teams are composed of officers who possess extensive experience in Military Classification and Assignment.

In an effort to control the reclassification [2] of these highly trained technicians, Headquarters, Army Air Forces, issued a Memorandum [3] making all reclassification subject to approval of a specially appointed Board of Officers (one of whom must be a Classification and Assignment Officer), at every Air Forces Air Field or similiar installation. Evidence must be established during the board meetings to the effect that the proposed reclassification of an incumbent will be of benefit to the Army Air Forces. Such reasons as: physically incapable of performing duties of the old job, proved inefficiency by actual on-the-job trial, or the incumbent's qualifications are needed to perform in a more critically needed job specialty, would justify reclassification.

---

* This Research Project was conducted during 1943 under the general supervision of Col. Alvin L. Prichard, Asst. Chief of Staff, A-1, Army Air Forces Eastern Flying Training Command.

** Capt. Arthur O. England is Officer in charge of Classification Research Section, Army Air Forces Eastern Flying Training Command.

[1] War Department Memorandum No. W615-9-43, Adjutant General's Department, January 22, 1943.

[2] Reclassification is a term used to describe any change in the classification of the incumbent. It may be downward (representing less utilization of the individual's capabilities), upward (representing greater utilization of the individual's capabilities) or lateral (representing less utilization of the individual's capabilities, but in a different field of skill). Insofar as we are concerned in this article, reclassification will be construed to mean downward reclassification.

[3] Army Air Forces Memorandum No. 35-16, January 22, 1943.

Although this established control medium has, to a large extent, prevented wholesale malassignment and inefficient personnel utilization, the evaluation of each individual case appearing before the Classification Board was dependent ultimately upon the local, subjective interpretations of the values of the military jobs by the Board.

Job specifications of Military Occupational Specialties are, by necessity, designed for flexible interpretations to provide a method whereby an individual, performing duties on various types of equipment, peculiar to each Air Force or Command, can qualify as a specialist in this one general job. For example, there is one set of job specifications for Airplane Mechanic. However, the Airplane Mechanic who has been trained on a four-engine Flying Fortress is definitely a higher skilled mechanic and can perform a greater variety of mechanical duties than can a mechanic who has been trained on a single engine Primary Trainer Airplane. Both, however, receive the same rating of Airplane Engine Mechanic.

It was because of the above instances of a lack of any standardized method of evaluating Army jobs, along with the relatively high degree of subjective interpretations that has been entering into the problem of reclassification and personnel utilization, that the publication "Job Evaluation for Army Military Occupational Specialties," [4] appeared December 18, 1943. The problem at hand was to provide a means of establishing standardized values for different Army jobs for the purpose of ascertaining efficient utilization of available manpower.[5]

Procedure: A Factor Analysis

An attempt was made to identify all of the possible factors for a job analysis. Nine significant items were finally selected and retained in the rating scale. Each item is divided into six component parts, with a numerical value assigned to each. The numerical value of each item de-

---

[4] Job Evaluation for Army Military Occupational Specialties, developed by Capt. A. O. England, Headquarters, AAF Eastern Flying Training Command, Classification Research Section, Maxwell Field, Alabama.

[5] It should be noted here that the publication of "Job Evaluation for Army Military Occupational Specialties" was intended to serve solely as a "guide" and does not represent an official publication by the War Department.

Table 1

Factor Analysis and Point Values on Job Rating Scale

Items	0	1	2	3	4	5
Educational Requirement	None	Grammar Sch.	High Sch. Grad.	Some College	College Grad.	Post Graduate
Civilian Experience	None	2 Months	6 Months	1 Year	3 Years	5 Years and Over
Training Time on-the-job	3 Wks. and Less	1 Month	3 Months	6 Months	9 Months	12 Months and Up
Army Tech. School (Length of course)	None	6 Wks. and Under	8–10 Wks.	12–14 Wks.	16–18 Wks.	20 Wks. and Up
Intellectual Requirement by AGCT Test Score (min.)	None	Below 69	70 to 89	90 to 109	110 to 129	130 and Up
Skill	None	Passable	Fair	Good	Excellent	Superior
Supervisory Capacity (No. men supervised)	0–4	5–9	10–19	20–39	40–79	80 and Up
Nature of Work	Routine	Passable Empl. of Ingenuity	Average Empl. of Ingenuity	Good Empl. of Ingenuity	Excellent Empl. of Ingenuity	Superior Empl. of Ingenuity
Equipment Used on Job	None	Simple: Hand Tools, Saw, Drill, etc.	Complex: Teletypewriter, Typewriter, Testing Set, etc.	Very Complex: Comptometer, Flexible Guns, Key Punch	Intricate: Radio, Power Turret	Very Intricate: Radar, Bombsight

pends upon its relative position in the table (Table 1), i.e., the level of difficulty, or the higher the requirements, the higher the point value assigned to each item. The point value range extends from 0 points to a maximum of 5 points. This range was arbitrarily agreed upon. It could have been 0 to 50, or 10 to 100, or any progression of numbers.

In the selection of differentiating items, one very significant item had to be discarded, namely wage scale. Unlike most civilian rating scales, pertaining to the job, which are usually weighted heavily, the Army does not adhere to any specific wage scale for military jobs. That is to say, a soldier's grade (Private, Sergeant, etc.) does not necessarily prohibit him from performing any military job. Because of a multitude of factors influencing the promotion of enlisted men, it has been impossible to correlate any job to a particular grade. There are numerous instances of Sergeants working for Corporals because the Corporals are more qualified by reason of proven experience on certain type aircraft than other soldiers of higher grade.

Items to be Rated

a. Educational Requirement: The range extends from none to Post Graduate. Even though exact information on educational requirements is obscure for any military specialty, the majority of specialties have some minimum requirement, either clearly established in AAF Regulation 50-12 or arrived at by a consensus among those individuals exercising supervisory jurisdiction, even the particular military specialties. Point values assigned to the educational levels are indicated in Table 1.

b. Civilian Experience: There are a few selected jobs in the Army that require civilian experience of varying degrees. Factors influencing the awarding of point values to this item are the existence or absence of similar training in the Army. Tabulating machine operators, office machine servicemen, and psychological assistants are examples of military specialties that require a civilian background in that specialty before assignment to the specialty is made. Point values are assigned according to the time spent gaining experience in that specialty in civilian life.

c. Training Time on the Job: A consensus among competent personnel was used to assign point values for the "average" time spent becoming skilled in the specialty.

d. Intellectual Requirement,—By Minimum AGCT Score:[5] Once again, in those instances where the AGCT Score requirements are not clearly prescribed by regulations, such as the AGCT requirements for attendance at Technical Schools, a consensus among competent personnel was obtained. Point values assigned to this item are broken down according to the five groupings of AGCT Scores.

[5] AGCT: Army General Classification Test.

e. Skill: This represents a combination of aptitude and acquired talent. Perhaps this is one of the weaker items in this factor analysis because of the difficulty involved in rating the "skill" employed in each military specialty. A point of issue may very justly be taken on the line of demarcation between "passable" and "fair" skill. The extremities of this range are relatively clear; Kitchen Police, requiring "no" skill, and IBM Machine Repairman, requiring "superior" skill, are cited as examples of the extremes of the point value range. It was only after constant comparison of all factors involved in the performance of each job that the ensuing ratings in the item were arrived at.

f. Supervisory Capacity,—Number of Men Supervised: The point values assigned to this item are not dependent upon the number of men one has jurisdiction over, but rather the number of men one directly supervises and is responsible for in the exact performance of his job.

g. Nature of Work: The range of this item is from "routine" to "superior employment of ingenuity." Here, creativeness, originality and independent responsibility are evaluated.

h. Equipment Used on the Job: No particular difficulty was encountered in placing the proper point values to the breakdown of the scale in this item. The range extends from none to very intricate.

The difficulty of all rating scales lies in the proportionate point value assigned to each component part of each item. Which items shall be weighted the most, which the least, will depend upon a general consensus among those people who are cognizant of the job requirements being considered. That does not necessarily represent an inherent weakness in Rating Scale theory, but rather a weakness in our knowledge of the requirements of Army jobs.

Factor Analysis of Military Specialties Authorized for Use in the Army Air Forces

In this section, a factor analysis of Military Occupational Specialties has been performed by items on the basis of Family Groupings of jobs. The total points awarded each Military Specialty have been rated by the rank difference method. The relative rated position each specialty holds to each other is indicated in the column entitled "Final Rated Positions." The jobs have been arranged in a series, with the range extending from the highest evaluated job to the lowest evaluated job.

Not only is the "Final Rated Position" important in the evaluation of a job, but also the actual total point difference between the evaluated jobs should be given consideration. This becomes at once apparent when comparisons are made among Military Specialties on Single, Twin and Four Engine Airplanes, and when interchange of jobs among different Family Groups is attempted.

Table 2

Job Evaluation Ratings of a Sample of Five Military Job Groups

SSN	Job Title	Educational Requirements	Civilian Experience	Training Time on the Job	Army Tech. School	Intellectual Requirement	Skill	Supervisory Capacity	Nature of Work	Equipment Used on the Job	Total Points	Final Rated Position
	Administrative and Clerical Group											
826	AAF Supply Technician	1	0	4	0	3	2	1	3	0	14	7
821	QM Supply Technician	1	0	4	0	3	2	1	2	0	13	8
405	Clerk, Typist	2	2	2	0	3	2	0	2	2	15	6
275	Classification Spec.	2	0	4	1	4	4	0	4	0	19	3.5
055	Clerk, Non-Typist	2	0	2	0	3	2	0	2	0	11	9
056	Postal Clerk	1	0	2	0	3	1	0	0	0	7	13
274	Public Relations Spec.	3	3	3	0	4	4	0	4	0	25	1
213	Stenographer	2	2	2	0	3	3	0	2	2	16	5
835	Supply Clerk	1	0	3	0	3	1	0	1	0	9	11
502	Administrative Spec.	2	0	4	0	4	4	2	4	2	22	2
623	Financial Typist Clerk	2	2	3	0	4	3	0	3	2	19	3.5
348	Parts Clerk, Auto.	1	0	3	0	3	1	0	2	0	10	10
667	Message Center Clerk	1	0	2	0	3	1	0	1	0	8	12
	Armament and Ordnance Group											
911	Airplane Armorer	1	0	3	2	3	3	0	2	3	17	4
612	Airplane Armorer-Gunner	1	0	4	3	3	3	0	2	3	19	7
511	Armorer	1	0	3	0	3	2	0	2	1	12	5
683	Bombsight Mechanic	1	0	4	5	4	5	0	5	5	29	1
901	Munitions Worker	1	0	2	0	2	0	0	0	0	5	6
678	Power Turret and Gunsight Mechanic	1	0	5	3	3	4	0	3	4	23	2

Table 2 (*Continued*)

SSN	Job Title	Educational Requirements	Civilian Experience	Training Time on the Job	Army Tech. School	Intellectual Requirement	Skill	Supervisory Capacity	Nature of Work	Equipment Used on the Job	Total Points	Final Rated Position
	Airplane Maintenance Group (Based Upon Four Engine Planes)											
689	Airplane Cable Mech.	0	0	2	0	3	1	0	1	1	8	15.5
685	AP Elec. Mechanic	1	0	4	5	3	3	0	3	2	21	3.5
528	AP Hydraulic Mechanic	1	0	4	1	3	3	0	3	2	17	9
686	AP Instrument Mechanic	1	0	4	5	3	3	0	3	2	21	3.5
750	AP Maint. Technician	1	0	5	4	3	5	5	5	4	32	1
684	AP Power Plant Mech.	1	0	4	5	3	3	0	3	1	20	5
687	AP Propeller Mech.	1	0	4	5	3	3	0	2	1	19	6.5
555	AP Sheet Metal Worker	0	3	3	0	3	2	0	2	1	14	11.5
550	Airplane Woodworker	1	0	3	0	3	2	0	2	1	12	13
747	AP and Engine Mech.	1	0	3	4	3	2	0	2	1	16	10
548	Fabric and Dope Mech.	0	0	2	0	3	2	0	2	1	10	14
748	AP Mechanic-Gunner	1	0	5	5	3	3	0	3	3	23	2
665	Fuel Cell Repairman	1	0	4	4	3	3	0	2	1	18	8
559	Glider Mechanic	1	0	3	2	3	2	0	2	1	14	11.5
114	Machinist	0	3	4	0	3	3	0	3	3	19	6.5
256	Welder, Combination	0	0	3	0	2	1	0	1	1	8	15.5
	Transportation Group											
014	Auto Equip. Mech.	1	3	2	0	2	2	0	2	1	13	1
345	Auto Equip. Operator	0	0	2	1	2	0	0	0	1	6	4
932	Special Vehicle Opr.	0	0	2	1	2	1	0	1	1	8	2
931	Heavy *Auto* Equip. Opr.	0	0	2	1	2	0	0	1	1	7	3
	Chemical Group											
870	Chemical Tech.	2	0	4	3	4	3	1	4	3	27	1
809	Decontamination Equip. Opr.	2	0	3	2	3	2	0	2	3	17	2
786	Toxic Gas Handler	1	0	2	2	2	2	0	1	3	13	3

In this article, only five Military Job Groups have been presented. Actually, all the existing pertinent Army Air Force jobs have been evaluated in their respective Family Groups. In addition to the distinct Family groupings of jobs, one other Group, known as the "Miscellaneous Group," was brought into play. This Group represents a heterogeneous grouping of military jobs. Actually, many different family groups are represented and because they are, it was impossible to differentiate them by the factor analysis method. Examples of jobs falling into this grouping are Bricklayer, Laundry Technician, Translator, Bandsman, Dog Trainer, etc. Interchange of duties within this miscellaneous group will be dependent upon the discretion of the Classification Board after a careful study has been accomplished on the proposed transfer.

Index for Evaluated Military Occupational Specialties

The evaluated jobs were prepared in a readily accessible list. This listing, by family groups, shows the different jobs by their degree of "over-all" difficulty. Reference may be made to the index of evaluated jobs under the following circumstances:

(1) To ascertain if one job is upgraded or downgraded relative to another in the same family group.

(2) To ascertain the correctness, from a classification standpoint, of proposed new duty assignments in relation to the present duty assignment.

(3) To ascertain the validity of reclassifying an enlisted man from one Military Occupational Specialty to another.

(4) To reallocate enlisted men efficiently within a Command.

The interchange of jobs from one family group to another has always presented a challenging problem to the Classification and Assignment Officers. However, until such time as an evaluating scale has been devised whereby family grouping jobs may be compared, it appears highly improbable that a valid comparison can be made between the total awarded points of a specialty in one family group to the total awarded points of a specialty in another family group. A specialty requiring a high degree of mechanical skill, thus being rated highly in the evaluated job scale, cannot be compared adequately, say to a particular clerical job requiring a high degree of creativeness.

Preparation of a Reallocation Adjustment Table

An attempt has been made to prepare a reallocation adjustment table to be used as a guide in the reallocation of enlisted men by the qualifications in a particular Military Occupational Specialty, between single, twin and four engine flying schools.

The Personnel Divisions of any Air Force headquarters have from

Table 3

Index for Evaluated Military Occupational Specialties: Final Listing of Rated Jobs by Level of their Difficulty

SSN	Job Title	Total *Points*	Final Rated Position
	Administrative and Clerical Group		
274	Public Relations Specialist	25	1
502	Administrative Specialist	22	2
275	Classification Specialist	19	3.5
623	Financial Typist-Clerk	19	3.5
213	Stenographer	16	5
405	Clerk-Typist	15	6
826	AAF Supply Technician	14	7
821	QM Supply Technician	13	8
055	Clerk, Non-Typist	11	9
348	Parts Clerk, Automotive	10	10
835	Supply Clerk	9	11
667	Message Center Clerk	8	12
056	Postal Clerk	7	13
	Armament and Ordnance Group		
683	Bombsight Mechanic	29	1
678	Power Turret and Gunsight Mechanic	23	2
612	Airplane Armorer-Gunner	19	3
911	Airplane Armorer	17	4
511	Armorer	12	5
901	Munitions Worker	5	6
	Chemical Group		
870	Chemical Technician	27	1
800	Decontamination Equipment Operator	17	2
786	Toxic Gas Handler	13	3
	Airplane Maintenance Group		
750	Airplane Maintenance Technician	32	1
748	Airplane Mechanic-Gunner	23	2
685	Airplane Electrical Mechanic	21	3.5
686	Airplane Instrument Mechanic	21	3.5
684	Airplane Power Plant Mechanic	20	5
687	Airplane Propeller Mechanic	19	6.5
114	Machinist	19	6.5
665	Fuel Cell Repairman	18	8
528	Airplane Hydraulic Mechanic	17	9
747	Airplane and Engine Mechanic	16	10
555	Airplane Sheet Metal Worker	14	11.5
559	Glider Mechanic	14	11.5
550	Airplane Woodworker	12	13
548	Fabric and Dope Mechanic	10	14
689	Airplane Cable Mechanic	8	15.5
256	Welder, Combination	8	15.5
	Transportation Group		
014	Auto Equipment Mechanic	13	3
932	Special Vehicle Operator	8	2
931	Heavy Auto Equipment Operator	7	3
345	Auto Equipment Operator	6	4

time to time experienced difficulty in the reallocation of highly trained military specialists, especially when this reallocation necessitated the involvement of personnel at single, twin and four engine flying schools. It has been far too great a task to expect Headquarters Assignment Officers to analyze each enlisted man's qualifications to ascertain the type of equipment on which the individual has had training, thus earning a skilled or semi-skilled rating in a Military Occupational Specialty.

Such problems as "can an Airplane Maintenance Technician, experienced only on single engine, Basic Training planes, be adequately employed as an Airplane Maintenance Technician on four engine Flying Fortresses" are confronted daily by Personnel sections. Yet, objective answers to such problems are not to be found in existing Army directives. Further, the Engineering Officers at different airfields have been faced also with this perplexing problem, namely, that enlisted men qualified as Airplane Maintenance Technicians on Basic Training planes, were assigned to them for work as Airplane Maintenance Technicians, yet the airfield only has four engine planes stationed there. In the interest of good military classification, these enlisted men have earned their classification as Airplane Maintenance Technicians and should not be deprived of it. Yet in the interest of efficient engineering operations, the men are not qualified to perform the duties of a Maintenance Technician on, let's say, Liberator Bombers, B-24 planes. Personnel shortages at different airfields are rated only by the general Military Occupational Specialties without any specific reference to the type of equipment or airplane on which the men have had experience. Further, personnel are reallocated from airfield to airfield simply by mass numbers, disregarding, because of necessity, experience levels or types of experience of the individual man involved.

In order to overcome some of the personnel reallocation problems outlined above, it was decided that a comparison of indices for evaluated military specialties relative to single, twin and four engine airplanes be conducted. To compare adequately the total evaluated points assigned to each military specialty at single, twin and four engine flying schools, a job analysis was conducted, at representative stations of each of the aforementioned flying schools. Only those Airplane Maintenance Military Specialties performed at all three types of flying schools were subjected to the factor analysis breakdown. Comparative figures among the different types of single engine planes, as well as among the twin and four engine planes, proved to be insignificant. Therefore, the reallocation adjustment table (Table 4) has only three main groupings—single, twin and four engine.

Reasignment of surplus specialists or the reassignment of malassigned specialists, for corrective measures, between different types of flying schools may be made efficiently if the relative evaluated job ratings at the different type of schools are made accessible to those in control of reassignment. For example, should an airplane Crew Chief on a Basic Training aircraft be sent to a four engine flying school if no vacancies exist at any other single engine schools? Reference to the Reallocation Adjustment Table presents the following picture:

Table 4

Reallocation Adjustment Table

Single Engine		Twin Engine		Four Engine		
Suffix I		Suffix II		Suffix IV		
SSN	Job Title	SSN	Job Title	SSN	Job Title	Total Points
				750	AP Maint. Tech.	32
		750	AP Maint. Tech.			31
						32
						29
						28
750	AP Maint. Tech.					27
						26
						25
						24
						23
						22
		685	AP Elec. Mech.	686	AP Inst. Mech.	21
				685	AP Elec. Mech.	21
		686	AP Inst. Mech.			20
686	AP Inst. Mech.	687	AP Prop. Mech.	687	AP Prop. Mech.	19
687	AP Prop. Mech.					18
685	AP Elec. Mech.					18
		528	AP Hydraulic Mech.	528	AP Hydraulic Mech.	17
528	AP Hydraulic Mech.			747	AP and Eng. Mech.	16
		747	AP and Eng. Mech.			15
747	AP and Eng. Mech.	555	AP Sheet Metal Wkr.		AP Sheet Metal Wkr.	14
555	AP Sheet Metal Wkr.					14

a. Airplane Maint. Technician at a Single Engine School has been *awarded 27 rated points.*

b. Airplane Maint. Technician at a Four Engine School has been awarded 32 rated points.

From the standpoint of effective placement of personnel, a plus or minus (±) two (2) evaluated points is considered a reasonable difference to overcome, without losing too much training time, in the adjustment of a specialist to a different type of job training. The adjustment problem involved here becomes, at once, apparent. This Airplane Maintenance Mechanic on the Single Engine plane would find it almost impossible to overcome the five (5) points difference were he to be placed on duty as an Airplane Maintenance Technician on a Four Engine plane. Thus, if he were transferred to a Four Engine Flying School, he would have to start in the lower category job, Airplane and Engine Mechanic (747).

Thus, it is apparent that reference must be made to the type of equipment upon which the enlisted man has won his rating as a Military Specialist and how the point value of that specialty at each type of Flying School compares before correct and efficient reallocation of military specialists can be effected.

Summary

In summarizing, the salient features of this job study have been:

1. The preparation of a standardized method to interpret the relative level of difficulty of different military jobs.

2. The preparation of point rating scale, with assigned values to each job, which is based upon nine significant job features. This helps reduce the subjective interpretation placed upon broad, generalized job specifications contained in the army publications.

3. The preparation of a reallocation table, applying the principles of factor analysis, to assist those military personnel people in the task of reallocating their overages and shortages of personnel most efficiently.

4. Although this job study was only utilized extensively in the Army Air Forces Eastern Flying Training Command, Headquarters, Army Air Forces, requested that distribution of "Job Evaluation for Army Military Occupational Specialties" be made for all Air Forces and Commands within the continental limits of the United States for their inspection and criticism. Reports received from these other Air Forces and Commands were most favorable and the general validity of the devised rating scale was not challenged.

5. The problem of military job classification, evaluation and personnel utilization is constantly assuming added significance. Perhaps some of the ideas set forth in this presentation will in some small way shed additional light upon these intricate problems.

Received October 17, 1944.

Aircraft Recognition: II. A Study of Prognostic Tests *

Lester Luborsky

Duke University

The present article reports an attempt to determine characteristics of the individual which influence his final level of performance in short-exposure aircraft recognition training. It was hoped that the approach followed would permit (*a*) the construction of a prognostic test battery, and (*b*) an analysis of the concomitants of the final level of performance. Because of the dearth of knowledge on this subject, the selection of factors to be studied was necessarily dictated by personal opinion as to the factors most likely to be involved, the possibilities of rapid measurement, and certain related aspects of importance in the life situation towards which the Navy recognition training program was directed. An instance of the last is speed of recognition.

The data in this paper were obtained during a previously reported experiment (2) in which 4 equated groups, totaling 30 subjects, were taught aircraft recognition in a standardized manner. The subjects were pre-aviation V-5 students at Duke University.

Before training, a battery of "pre-tests" was administered which was intended to (*a*) measure possible perceptual and learning factors associated with "aircraft recognition ability" and (*b*) aid in equating the groups. After training, an extensive battery of "post-tests" was administered to facilitate analysis of the final level of performance. It is the scores of all subjects on these pre- and post-tests and on the Final Test of aircraft recognition which comprise the main data of this paper.

In accomplishing the first aim, the selection of tests which have predictive value for the outcome of training, the procedure essentially was to correlate scores of all 30 subjects on the pre-tests with scores on the criterion test, in this case, the Final Test. The tests which correlated significantly were treated by the Doolittle technique to obtain the multiple correlation with the Final Test and the percentage of the variance accounted for by each test. A further subsidiary aim was to obtain at least suggestive evidence as to the concomitants of the final level of performance in aircraft recognition. This was based upon the intercorrelations of all tests, pre- and post-, and the necessarily subjective estimate of what each test tests. Because of the relatively apparent meaning of some of the tests, such as A. C. E. Psychological Examination and the Test of Previous Knowledge, and the large number of test intercorrelations, the task of interpretation was considerably facilitated.

---

* Part of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences of Duke University, 1945.

Procedure

Classes were held for 45 minutes on Monday, Wednesday and Friday for a 6 week period. The pre-tests listed in Table 1 were administered on two successive evenings 3 days before the training period began. The "Pre 1" tests were given on the first evening and the "Pre 2" tests, which were alternate forms of Pre 1 tests, were given on the second evening. The post-tests were on the Monday and Wednesday classes of the last week. These tests also were alternate forms with the exception of

Table 1

Correlations of all Data with the Criterion
Pre 1, Pre 2, and Post Refer to the Time of Administration

*Span of Apprehension**		*Interest I*	.53
Pre 1	.13	*Interest II* (Improvement)	.47
Pre 2	.05	*Previous Knowledge of Planes*	
Post	.04	Pre 1	.41
Improvement Pre 1—Post	−.13	*Reaction Time*	
Improvement Pre 2—Post	.05	Light	
Memory Location		Pre 1	.02
Pre 1	.11	Post	.03
Pre 2	.10	Common Objects	
Post	.18	Pre 1	.18
Memory-Complex Figures, Part 1		Post	.41
Pre 1	.38	Improvement Pre 1—Post	−.06
Memory-Complex Figures, Part 2		Recognition of Planes	
Pre 1	.20	Post	.26
Memory-Airplane Recognition Aspects		*A.C.E. Psychological Examination*	.15
Pre 1	.29	*Grades in Academic Work*	.14
Post	.20	*Study Time on Aircraft Recognition 3rd Week*	−.57
Memory-Complex Figures, 5"			
Pre 1	.03	*Study Time on Aircraft Recognition 4th Week*	−.43
Acuity			
Pre 1	.03		
Post	.36		
Improvement Pre 1—Post	.54		

---

* Correlation coefficients after scoring by an alternative method (amt. ±) are: Pre 1 = .16; Pre 2 = .19, and Post = .06. The consistent lack of correlation is important because this test has been used in some recognition schools before training on the assumption that it improved aircraft recognition ability.

Acuity and Recognition Time for Planes. In addition to the above, A. C. E. Psychological Examination scores and Grades in Academic Work were obtained from the Dean's office and Aircraft Recognition Study Time was obtained from a questionnaire on study habits.

The experimenter operated a Balopticon with a shutter attachment for obtaining 1/50″ to 1″ exposures (by calibration, 1/40″ and 4/5″ respectively). The stimulus figures were projected on a screen about 5 feet from the subjects. The testing was done in groups no larger than 8 to lessen the possibility of differences between subjects caused by seating position.[1] The less essential details of the apparatus and testing conditions are more fully described in another article (2).

Description of Tests [2]

1. *Span of Apprehension* (*Pre 1*, *Pre 2*, *Post*). The subjects were required to estimate the number of 5 centimeter [3] randomly arranged horizontal lines flashed on the screen for 1/50″. The projection area of the screen was approximately 15″ × 18″. Each test contained 16 exposures with 4, 6, 8, or 10 lines in each. The scoring was done in two ways: (*a*) the number of incorrect estimates, and (*b*) the sum of the deviations from the actual number of lines exposed.

2. *Memory-Location* (*Pre 1*, *Pre 2*, *Post*). A test of ability to reproduce on a grid sheet the location of a group of 3⅛″ × 5″ rectangles flashed on the screen for 1/50″. Each test contained 12 exposures of 5, 6, or 7 rectangles each. The score for each grid was the total of the omitted and incorrectly filled-in grid spaces.

3. *Memory-Complex Figures, Part 1, 1/50″* (*Pre 1*). Subjects must reproduce simple geometric line figures presented for 1/50″. Only 1 exposure was given for each of the 4 figures constituting the test. An arbitrarily-devised scoring system credited one point for each of 4 aspects of each figure.

4. *Memory-Complex Figures, Part 2, 1/50″* (*Pre 1*). Subjects must reproduce geometric line figures of greater complexity than in Part 1, presented for 1/50″. Each of 3 figures was given 4 times in succession and after each presentation the earlier reproductions were covered with the left hand while the new reproduction was drawn. The scoring was 1 point for each of 6 arbitrarily chosen aspects of each figure. The points for each reproduction were totaled to yield a score for each figure.

5. *Memory-Airplane Recognition Aspects, 1/50″* (*Pre 1*, *Post*). Plan view diagrams of airplanes, taken from *Aeronautics Aircraft Spotters' Handbook* (1) were presented for 1/50″. Sixty-five seconds were allowed after each exposure of a plane to check the presence or absence of 14 recognition aspects on a mimeographed check list. Subjects were informed that these were planes which were infrequently used and hence unfamiliar to them (as, e.g., Loire-Oliver 45, Falcho, Morane-Saunier 406). Before starting the test every effort was made to be sure that the recognition aspects in the check list were completely understood. To equalize the amount of guessing, subjects were directed

---

[1] Analysis showed no differences due to seating position. Staff, Psychological Test Film Unit (3), mention like conclusions.

[2] Fuller descriptions and illustrations of the test materials are presented in the doctoral dissertation available through the Duke University Library.

[3] Five centimeters after projection on the screen. The increase in size on projection is 1 to 5.

to leave no omissions. The Pre 1 test contained 10 planes, and Post test contained 5 additional planes.

6. *Memory-Complex Figures, 5″ (Pre 1, Pre 2)*. This test is similar to the Memory for Designs test in the Stanford-Binet. The 2 cards in Pre 1 were almost identical with the test cards in Form L and M of the Stanford-Binet. Each test card contained 2 figures which were presented simultaneously for 5″. After approximately 3 seconds subjects were reminded to look at both figures. The figures on the Pre 2 cards[4] were considerably more complex than those of Pre 1. A more refined scoring than that used by the Stanford-Binet was devised—1 point for each of 5 aspects of the first figure and 1 point for each of 3 aspects in the second figure of *each* card.

7. *Acuity (Pre 1, Post)*. This test employed the Snellen Symbol E Chart.[5] The subject indicated the direction of the open end of the E using two-eyed vision. The score was the number of E's correctly read, on a chart with 38 letters.

8. *Interest I*, and 9. *Interest II*. These tests were given at the end of the fourth week of training. By way of introduction, the experimenter stated that some planes would be described which would not be included in the course. However, students might be interested in these planes since they were commonly used in training aviation personnel and in passenger service. A test was given of 6 planes, 3 views for each, to see how many were already known before teaching. This test will be referred to as Interest I. The names, recognition features, and interest items for each plane were then discussed while viewing each plane for approximately 2 minutes each. Immediately after this teaching and without any previous warning another test was given to see how much had been learned. The Interest II scores are the improvement from the first to the second test. The scores on both the Interest I test and the Interest II test are considered as interest measures—the former on the assumption that those students interested in aviation would have known more of the common planes, and the latter on the assumption that those students with greater interest would be more receptive to the teaching.

10. *Previous Knowledge of Planes (Pre 1)*. This test contained 33 of the 36 planes to be included in the syllabus, given a 3″ exposure for each. All were plan view diagrams.

11. *Reaction Time*. Before and after training, reaction times were taken for (*a*) seeing a 1/50″ light flash on the projection screen, and (*b*) naming common objects, such as umbrella, book, cat, etc., flashed on the projection screen for 1/50″. These two kinds of reaction time were expected to prove useful when the recognition time for planes, measured at the end of training, was to be interpreted. The subjects were required to lift their forefinger from a stimulus key in circuit with a standard timer to indicate recognition and then name the plane.[6]

12. *American Council on Education (A. C. E.) Psychological Examination*. Scores on this test (total score on linguistic and quantitative sections) were obtained from the Dean's office. Most of the tests had been administered during the previous semester.

13. *Grades in Academic Work*. The Dean's office also supplied the previous semester's grades in academic work. Scores are the average of point equivalents for each grade: 13 points for a grade of A, 10 for B, 7 for C, etc.

---

[4] It was found that this set of stimulus cards correlated highly with Pre 1 but with no other test. The greater complexity of these figures probably introduced an additional influence to that measured in Pre 1.

[5] American Optical Co., No. 1042.

[6] The writer is indebted to Mr. Walter Knight of the Department of Physics, Duke University, for constructing the efficient reaction time shutter attachment.

14. *Study Time on Aircraft Recognition, 3rd Week.*

15. *Study Time on Aircraft Recognition, 4th Week.* The time spent in studying aircraft recognition outside of class was recorded on a questionnaire on study habits. The total time in minutes for each week was used for correlation purposes.

The Criterion Measure: the Final Test

All tests were correlated with the 1/50″ Final Test which was administered at the eighteenth session, the last session of the course. In the last 2 weeks of this course all groups were trained at the 1/50″ standard and given many views of each plane. One picture and one diagram view of all 36 planes taught gave a total of 72 exposures in the 1/50″ Final Test. This test was longer and more difficult than any previously given. The split-half reliability coefficient was +.77.

Results and Discussion

Tests with Significant Correlations with the Final Test and with Each Other. The correlation coefficients of all tests with the criterion are presented in Table 1. Table 2 presents only those pre-tests with signifi-

Table 2

Main Tests: Correlation with Criterion, Reliability, and Significant Correlations with Other Tests

	r with Final Test	Reliability: Self-r or Test-retest	Significant r's with Other Tests	
1. Memory-Complex Figures, Part 1, 1/50″	.38	Part 1 Part 2 } .43	Memory-Complex Figures, 5″	.36
2. Memory-Complex Figures, 5″	.63	Pre 1 Pre 2 } .74	Memory-Location, Pre 2 A.C.E. Psychological Examination Memory-Complex Figures, Part I	.54 .43 .36
3. Previous Knowledge of Planes used in the Syllabus	.41	.74	Interest I	.43
4. Interest I	.53	.42	Previous Knowledge	.43
5. Interest II	.47	—*	A.C.E. Psychological Examination	.64

* No satisfactory method is available for finding the reliability of this improvement score. However, an improvement score cannot be reliable if the tests from which it is calculated are unreliable. The split-half reliability of the test from which Interest II is derived are .42 and .52 respectively.

cant[7] correlations and also includes their reliability coefficients and a list of other pre- and post-tests with which these correlate significantly. The highest correlations with the criterion[8] (Table II) are (1) Memory-Complex Figures, Part 1; (2) Memory-Complex Figures, 5″; (3) Previous Knowledge of Planes used in the Syllabus; (4) Interest I, Knowledge of Planes (trainers and passenger); (5) Interest II, Improvement in Knowledge of Planes (the same planes as in Interest I). The significant correlations of these "main" tests with other tests should give some hints as to the nature of these main tests and at the same time a better understanding of the nature of "aircraft recognition ability." This problem will be more profitably considered after arriving at an estimate of the variance contributed by each main test to the total variance.

Development of a Prognostic Test Battery. In the development of a prognostic test battery, the intercorrelations of 5 main tests and the Final Test were treated by the Doolittle technique and a multiple correlation (R) of .855 obtained. The index of forecasting efficiency is 48.2%. The standard error of estimate was found to be 4.09. This means that in 67% of cases the predicted criterion score will be within 4.09 points of the actual criterion score. The R has a coefficient of multiple determination of approximately 73% of the variance in the Final Test. Each variable contributes the following amount to the total variance: X_2, Memory-Complex Figures, Part 1, 1/50″, 8.0%; X_3, Memory-Complex Figures, 5″, 28.9%; X_4, Previous Knowledge of Planes, 9.5%; X_5, Interest I, Knowledge of Planes, 13.0%; and X_6, Interest II, Improvement in Knowledge of Planes, 13.6%. By far the greatest contribution is made by Memory-Complex Figures, 5″. Interest I had Interest II make the next greatest contributions. However, all contributions are large enough to make it desirable to use all tests if this battery were to be used prognostically.

Beta weights were then computed and the following regression equation for predicting the criterion scores from the 5 tests was found: $X_{1\,\text{Pred.}} = -3.100 + 1.033\ X_2 + 1.815\ X_3 + 0.443\ X_4 + 0.576\ X_5 + 0.117\ X_6$.

It was suggested above that all of the tests should be used for prognosis since none of the variance contributions is small enough to disregard.

---

[7] All correlations must be .36 to be significant at the 5% level and .46 to be significant at the 1% level.

[8] Since the highest obtained correlation coefficient between consecutive aircraft recognition tests is only .64, a correction for attenuation would considerably increase all correlations. However, it was thought more desirable in this preliminary study to present all the coefficients exactly as obtained and merely note the fact that a correction for attenuation would have the effect of increasing all coefficients by $1/\sqrt{.644}$ or e.g. if $r = .38$, $r_c = .48$.

Use of all tests presents no administrative difficulties, since all tests would take less than 45 minutes.[9] However, after this battery is validated on another new and larger group, it may be found that inclusion of some of the tests is not economical. Furthermore, it is likely that lengthening of certain of the tests will result in an even greater R. The high R already obtained indicates that the preliminary search for prognostic tests appears to have succeeded in sampling the major areas contributing to success in aircraft recognition training.

Table 3

Diagrammatic Analysis of Interrelationships (Significant Correlations) Between the Main Tests;* Final Test and the Main Tests; and the Main Tests and Other Tests

Kind of Factor	Main Tests (Correlated Significantly with Final Test)	Tests Correlated Significantly with Main Test but not with Final Test
	Measured before Training:	
Memory	Memory-Complex Figures, Part 1, 1/50″	
	Memory-Complex Figures, 5″	Memory-Location A.C.E.
Past Knowledge	Previous Knowledge of Planes	
	Interest I	
Interest	Interest II	A.C.E.
	Measured after Training:	
Acuity Improv.	*Improv. in Acuity*	
Reaction Time	Reaction Time: Common Objects (Post)	Reaction Time: Recognition of Planes (Post)
Study	Study Time for Aircraft Recognition**	

* Main tests which are significantly correlated are bracketed.

** A negative correlation.

Analysis of Conditions Influencing Performance. One aim of this experiment was to obtain at least suggestive evidence which would increase understanding of the concomitants of successful performance. The variance contribution of each pre-test to the total variance is presented in the preceding section. In the present section a more complete analysis will be attempted. For this purpose all tests, rather than only the pre-tests, with significant correlations with the criterion, must be considered. In addition, these tests might be better understood by considering the tests with which they, in turn, correlate significantly. The diagrammatic analysis in Table 3 is based upon both sets of correlations.

---

[9] Administrative time required for each test: X_2 = five minutes; X_3 = five minutes; X_4 = ten minutes; X_5 = five minutes; and X_6 = twenty minutes. Total administration time = *forty-five minutes.*

A hypothetical analysis of the concomitants of successful performance based upon Table 3 might be as follows:

(*a*) Memory for complex figures, perhaps of a visual memory type. Both memory at 5″ and 1/50″ exposure are involved but particularly the former.

(*b*) Past interest in learning planes and resultant knowledge of some planes.

(*c*) Interest and intelligence[10] which permits ready learning of planes.

(*d*) A tendency to adapt to the visual conditions demanded by the mode of presentation of planes. This is an inference from the significant correlation of Improvement in Acuity and the Final Test.

(*e*) After training there is ability to recognize and name common objects more quickly under short exposure presentation conditions.

(*f*) Existence of the above make a great deal of studying unnecessary. Another possible interpretation is that much study of the kind done actually interfers with proper learning.

It is possible to infer from the variance contributions previously determined that the first 3 concomitants listed above are important for successful performance in the order listed. The factor which contributes the most to the final level of performance is the ability to remember complex material, relatively independently of exposure time. The past knowledge factor which contributes the second largest amount may also be considered a memory factor.

Summary and Conclusions

The data for the present report were obtained almost entirely from tests given before and after an aircraft recognition training course. A prognostic battery and regression equation were found by applying the Doolittle technique to pre-tests with significant correlations with the Final Test. All of these selected pre-tests were found to contribute sufficiently to the total variance to warrant inclusion in the prognostic battery. Because of the high multiple correlation of .85 and the ease of administration of these tests, validation of the battery on a new and larger group of students is recommended.

---

[10] Although the intelligence measure (A. C. E.) did not correlate significantly with the Final Test, it did correlate highly with Interest II.

On the basis of the significant correlations between individual tests and the criterion, a hypothetical analysis was attempted on the concomitants of achieving successful performance in recognition under conditions of short exposure. These concomitants are briefly as follows: (*a*) memory for figures; (*b*) previous knowledge of planes; (*c*) interest and ability in learning planes; (*d*) adapting to the short exposure visual conditions; (*e*) quick recognition of common objects after short exposure training; and (*f*) comparatively less home study.

These relationships should hold particularly for recognition of this specific stimulus material. However, there is good reason for supposing that the relationships would hold for recognition of material of comparable level of complexity and of similarity within the group of items.

Furthermore, the data presented in this section are of great utility for future experiments. The analysis of concomitants of achieving the final level of performance would enable a more complete equation of groups than was possible before this experiment, and would open the way to final answers to the kinds of problems described in an earlier experiment (2) on the relative efficiency of training procedures.

Received October 18, 1944.

References

1. Guthman, L. C. (Ed.). *Aeronautics aircraft spotters' handbook.* (4th Ed.) New York: National Aeronautics Council, Inc., 1943.
2. Luborsky, L. Aircraft recognition: I. The relative efficiency of teaching procedures. *J. appl. Psychol.*, 1945, 385–398.
3. Staff, Psychological Test Film Unit. History, organization, and research activities, Psychological Test Film Unit, Army Air Forces. *Psychol. Bull.*, 1944, 41, 457–468.

Psychological Principles in Army Administration

Louis L. McQuitty,* Lt. Col., AGD

University Training Command, MTOUSA, APO 49, c/o Postmaster,
New York, N. Y.

The army encompasses a tremendous administrative responsibility. This is true both from the point of view of the number of individuals administered and from the many details of human behavior minutely affected. Some military psychologists have proven so successful in their approaches to administrative problems that their commanders have placed them in larger administrative duties involving fascinating and practical psychological problems. Most of the psychological problems posed cannot be referred for final solutions to established laws, principles or axioms. An alternative approach is to obtain as much insight as possible through discussions by those trained in fields of knowledge closely related to the problems. This paper represents one such discussion. It outlines tried methods of solutions to army administrative problems and suggests the psychological principles from which they appear to derive.

Administrative Organization

One of the first problems containing psychological implications with which the army administrator must concern himself is the type of personnel organization he is going to install in order to accomplish the mission for which he is responsible. The size of the mission is often so great that the administrator cannot accomplish the job by himself and cannot even keep himself informed on all of its many aspects. It is therefore incumbent on the administrator that he divide the mission into goals and assign the goals to sub-administrators.

The division into goals should be based on a careful study of the mission. The mission requires execution of certain work functions in order to accomplish it. These work functions form clusters in the sense that the component functions of any one cluster are highly dependent upon one another but are relatively independent of the functions of other clusters. The division of the mission into goals should be so accomplished that each goal represents a cluster of work functions. An illustration of this point is readily available in the consideration of personnel administration and classification responsibilities as a mission. One possible division would be to have a personnel administration section and a classification section. Each section would handle its particular responsibilities as they pertained to both officers and enlisted men. An alternative division would be to have an officer section and an enlisted section. Each section would handle personnel administration and classification functions, one section as they pertain to officers and the other as they pertain to enlisted men. A comparative study of these two alternative divisions reveals that soldiers on duty in the officer section could operate satisfactorily with little reference to soldiers on duty in the enlisted section. The work functions represented by the officer and enlisted sections constitute clusters and the clusters are relatively independent. This condition is just the opposite of that which exists when the division is into a personnel administration section and a classification section. The soldiers on duty in these two sections must be continually communicating with one another in order to function satisfactorily. These two sections do not represent independent clusters.

---

* On military leave of absence from the Department of Psychology, University of Illinois.

The division of the mission into goals represented by relatively independent work clusters, as purposed herein, is considered to be desirable because of the following reasons:

(1) It minimizes the amount of coordination needed; (2) It lessens the chances for personality frictions which are apt to occur when persons of equivalent rank and responsibilities have to coordinate and determine for which aspect of a mission they are responsible; and (3) Each sub-administrator has the maximum amount of autonomy possible within army organization. Practically all his dealings are with his immediate superior or his immediate subordinates.

The administrator, unless he is charged with a small and minor mission, will find it advisable to delegate to sub-administrators all of the duties contained in the mission. This will leave the administrator free to discharge the responsibilities of general supervision, coordination with those on an equivalent administrative level, and to report to and receive directives from his immediate superior. General supervision should include: (1) A review of non-routine work submitted to higher authority; (2) A review of the assignments of incoming non-routine tasks to sub-administrators in order to assure that the original division into sub-missions is followed or revised when appropriate; (3) A continual review of personnel allotments to the sub-divisions in order to maintain equivalent work loads for all concerned as the amount of work charged to the sub-divisions fluctuates; and (4) Inspections of the accomplishments of sub-administrators.

If the administrator follows this plan of office organization, he will usually discover that from four to six is the appropriate number of sub-administrators for his immediate supervision.

Having attempted to so organize the functional aspects of the office to assure wholesome relationships between personalities, the administrator can further the realization of the desired ends by an appropriate distribution of office space. This statement is based on the assumption that desired cooperation is more likely to ensue if it is made easy, the parties concerned are friends, and each appreciates the problems of the others. It is believed that these latter conditions are usually achieved if the sub-administrators are placed in the same office or in adjoining offices and communication between them is readily available at all times.

It is felt that the above suggested organization tends to facilitate smooth operation because it decreases the chances for personality frictions by lessening the possibilities for misunderstandings and frustrations. It charges the administrator primarily with the job of making the organization function properly. This enables him to forsee possible personality difficulties before they arise and to act to correct them before they become serious.

Having organized the office, both from the point of view of space assignments and functional responsibilities, there are the problems of the relationships of the administrator to his sub-administrators, to his co-administrators and to his superior.

Relationships between Administrator and Superior

The significance of the relationship of the administrator to his superior is realized in part when one recalls that in army life a soldier is responsible to his superior for practically everything he does or fails to do. It is therefore incumbent upon a soldier administrator to build a wholesome relationship between himself and his superior. In order to achieve this end he should learn the details of his superior's personality. He should listen very attentively to everything he says. He should study his motivations and attempt to predict his reactions to situations as they arise. He should put tactful questions when answers to these assist him to know better the desires of his superior as they pertain to military matters and when the questions can be put without arousing offense. These suggestions derive from the principle that there are wide individual differences in superiors and what might appeal to one would be acted on unfavorably by another. The solution is to know the superior to whom one is responsible as well as possible so that appropriate matters may be presented in a manner so as to be favorably received.

Despite the individual differences in superiors there are usually some principles of approach that will be favorably received by most of them.

It is well to be acquainted with these because in army life there is often a change in the individual to whom any one is responsible, and the generally acceptable principles have to be relied upon until one can learn the *personality characteristics of his new superior.* One can try the generally applicable principles and observe closely their effects. He can revise them in the light of his observations.

Primary among the generally acceptable principles is the desirability of putting one's suggestions for the accomplishment of tasks or goals too his superior in such a way that they imply that the administrator fully realizes that decisions on the matters are the prerogative of the superior. If any of the suggestions are adopted and prove successful, the administrator can usually motivate wholesome relationships, if he makes inherent in his report of the success to the superior the realization that it was the judgment of the superior which decided in advance that the ideas would *be successful. This tends to encourage in the superior the feeling of* ownership for the program involved and thus the program obtains his support more effectively. Several repetitions of this nature create in the superior more and more of a feeling of personal ownership for all of the programs that derive from the administrator under consideration to the end that the latter obtains a very high level of support.

On the other hand, if an idea suggested by the administrator and approved by the superior should fail, it is well that the report of failure include the realization by the administrator that he suggested the idea and accepts responsibility for its failure. The necessary derogatory impression ensuing from a report of failure is lessened by including a suggestion—particularly well thought out—to improve or substitute for the idea which failed on first trial.

Many of the suggestions which the administrator recommends to his superior will originate with his immediate sub-administrators or even lower in the administrative echelon. It is well that the administrator credit them to the appropriate individuals when he recommends them to his superior because this is one way in which the administrator can encourage his superior to the realization that his actions are not predicated on a mere desire to create a particularly favorable personal impression. In addition, to give the credit where due will help one's relationship with one's sub-administrators because the chances are the superior will acknowledge the contribution at some time to the sub-administrators concerned, and the sub-administrators will feel kindly to know that they have been given due credit.

At times the administrator will receive from his superior directives with which he may immediately disagree. It is wise that he be hesitant about expressing his disagreement unless invited to do so, and then it should be done with reservations. It may be that the superior, and even his superiors, have spent a great deal of time thinking through and studying the directives. The directives may be the product of tremendous efforts. For the administrator, a junior, to object with little apparent study may not portray due respect and may be unfavorably received because of the apparent failure of the administrator to give due study to the directives before voicing his attitude toward them. Also if he takes time for himself and his sub-administrators to study them and report to him he may arrive at more valuable conclusions, which will be more respected because they have first been given due consideration, and the superior may not be as emotionally interested in the directives as he was soon after he may have spent hours working on them and thinking about them. Also the administrator establishes an attitude of confidence toward himself if he is very cautious about raising problems in connection with directives to him. He may be able to establish in his superior the attitude that the latter should consider very carefully when the administrator raises a problem. Such, of course, would be a definite asset to the administrator and would probably spread to become a more general attitude of regard by the superior toward the administrator.

Another possible way to build a favorable attitude in the mind of the superior and to save him embarrassment as well as put him in a position to justly defend the actions of his subordinates is for the administrator to keep the superior especially well informed on what transpires. This applies in particular when the administrator or his staff has been guilty of an oversight or some other type of negligence or when occasion has required the issuance to commanders of lower echelons directives to which they are expected to object. The chances are the negligence or objection will be reported through command channels and will come to the superior from personnel to which he is responsible. He can do much to retain cooperation and understanding among all concerned if he is in a position to give immediately the full story on the matter under consideration.

There is of course always a question of the minuteness of matters of potential objection and negligence which should be reported to the superior by the administrator. The answer to this problem depends on the personal preference of the superior and can usually be discovered by appropriate questioning when such matters are reported. His preference will usually reflect the attitude of the individuals to whom he is responsible and the tendencies of lower and higher commanders in reporting objectionable matters. Some commanders will attempt to report their objections with reasons to the office which issued the directive involved and which appears to be immediately responsible so that a more satisfactory arrangement may be evolved at that level if possible. Others will invariably report their objections to personnel in higher echelons of administration and therefore make it desirable for those personnel to keep well informed on the matters because they can thereby make a more favorable impression on the objecting commander, who usually ranks the administrative personnel, to the end that cooperative understanding has a better chance of continuing.

The points just outlined covering principles that are considered generally helpful to an administrator in adapting to his superior are predicated on the belief that personal understanding between the two concerned is of primary importance, that the administrator as the junior should consider it his responsibility to make whatever adaptation is necessary to accomplish the desired relationship, that the administrator should regard himself as an assistant to the superior in the accomplishment of the latter's mission, that frankness and cooperation are essential to personal understanding, and that credit will accrue to those to whom it is due much more readily if they make no obvious moves designed merely to enchance it—and besides—self credit, especially in a war effort, is of no consequence; the one great dedication should be successful and efficient prosecution of the war.

Relationships between Administrator and Sub-Administrators

The fundamental fact to keep in mind in considering the relationships of the administrator to his sub-administrators is that the discussion is here pitched at that echelon where the size of the mission is so large that its accomplishment has been entirely delegated—leaving the administrator free to give his full time to the supervision of his subordinates. The administrator is dependent upon his subordinates for accomplishment of the mission, and his success depends on his qualities of leadership in a particular field. His job is to assure that the maximum achievements are realized by his subordinates. These achievements must be realized as a functional component of a much larger whole. They must be appropriately adjusted to the much larger functional framework. The administrator must therefore be a successful teacher. He must be able to give his sub-administrators a clear concise picture of the mission for which he is responsible. He must be able to give a significant and meaningful picture of how this mission fits into the larger framework. He must be able to teach each sub-administrator *to realize and comprehend his* particular goal in its perspective to the goals of other sub-administrators and in perspective to the still larger functional framework. He can test his success. If his subordinates are making recommendations which his knowledge of the overall situation frequently reveals to him are inappropriate, then he is not keeping his subordinates sufficiently well informed—and if he is passing these recommendations on approved to higher administrative echelons where they are not favorably considered, then he is not keeping himself nor his subordinates sufficiently well informed. The suggestions just offered are believed to derive from the principle that appropriate achievements depend on clear cut understandings of missions and that one characteristic of respected leadership depends on an unbiased ability to search oneself and one's organization for shortcomings before blaming one's subordinates.

If the shortcomings are discovered in oneself, one must take immediate steps to correct them, and one way that one can usually assure himself that he and his subordinates are sufficiently well informed to properly discharge their duties is to attempt to keep himself sufficiently well informed to accomplish the mission of his superior and frankly attempt to keep his immediate subordinates sufficiently well informed that any one of them could step into his position. This has the effect of keeping a personal goal immediately in front of all personnel concerned and acts as a strong motivating factor. It also makes a very favorable impression on superiors to observe the continued smooth functioning that ensues even though key personnel are suddenly lost as so frequently happens in many army organizations.

The above suggestions encourage successful leadership through careful instructions to subordinates. Another method of giving direction to the work of subordinate is for the administrator as the leader to assist his subordinates to know his personality characteristics. The better able he is to acquaint them with his personality characteristics the better able they will be to follow his desires. He will assist his subordinates in this respect if he always gives reasons for his decisions. In addition, this policy makes him very careful of his decisions and often prevents him from making errors of decision, especially so, if he creates in his subordinates a feeling of freedom to express disagreements. This feeling is enhanced if such expressions are always given careful and tactful consideration and if appreciation is shown for them even though they are not accepted. The approach just offered is believed to derive from the principle that freedom of action—herein fostered by an overall knowledge of the mission plus responsibility for definite goals—encourages initiative.

Initiative can often be still further encouraged by emphasizing responsibility to personnel to whom goals are assigned. Such personnel will sometimes come to the administrator with questions designed to obtain the solution from him—and for the administrator to give it will encourage dependence upon him at the expense of properly developing his subordinates.

The assignment of goals also obligates the administrator from the should at least give them an opportunity to express their opinions on matters prior to a decision having been made. To comply with this obligation should not only assist in developing and maintaining desired personal relationships but should also produce valuable ideas which might otherwise be overlooked.

The suggestions enumerated above are based on the principle that smoother personal relationships are maintained and more is achieved when all personnel are given a clear cut understanding of their responsibilities, are taught to exercise their initiative, and the initiative of all is given proper direction by means of keeping all personnel sufficiently well informed so that they have proper perspectives for expressing themselves.

Relationships between Administrator and Co-Administrators

By keeping oneself especially well informed and by thinking out ones decisions with ones sub-administrators, as just outlined, the administrator keeps himself peculiarly adept to deal effectively with his co-administrators, with whom he must cooperate and with whom he is often in competition. He must cooperate with them because all are working for the accomplishment of the mission of their common superior and for the greater goal of successful and efficient prosecution of the war. He is often in competition with them because all are often highly motivated by the superior, all are rated as to efficiencies by their superior, and disagreements arise as to how the mission of the superior should be accomplished and as to whom of the sub-administrators should be responsible for the various aspects of the mission. *The administrator who is well reinforced* by knowledge and carefully thought out principles has the advantage because he is prepared for discussion periods which usually arise suddenly. *In alternative solutions that arise in these discussions* he should take only efficient accomplishment of the mission of his superior as the primary purpose, and this principle in conjunction with an unusual background of pertinent knowledge will give him a solution for which he can acquire support through an intellectual, non-emotional, presentation. This presentation will encourage better personal relationships if the intellectual contributions are offered in the form of suggestions and if the contributor takes advantage of every opportunity to evolve them from the personnel in the discussion. He will then usually be admired by his colleagues and appreciated by his superior.

The proposals herein *offered for efficient functioning of an administrative* organization are analogous to those often proposed for the healthy minded functioning of a human organism. They derive from the primary principle that efficiency will ensue in the *organization, or organism,* if there is a clear cut goal adjusted to the capacity present and if all components are carefully directed and motivated toward its realization.

Resumé and Psychological Possibilities

This paper was motivated by the realization that never before have psychologists contributed so much in so many assignments heretofore not filled by personnel especially trained in the understanding of behavior. The exigencies of war and the success of some military psychologists in non-specialized aspects of their assignments have placed them in these pursuits. This unusual situation demands discussion in order that it may be properly evaluated, in order that full advantage may be taken of its opportunities, and—more specifically—in order that psychologists in these assignments may take advantage of the contributions of their colleagues.

The paper outlines psychological approaches that have been applied to certain army administrative problems. The discussion is pitched at that level where the administrator must depend entirely on his subordinates for the accomplishment of the tasks. His achievement depends on leadership and organization. He teaches his subordinates to understand clearly their goals in proper perspective to the larger missions. He encourages and developes responsibility, initiative and determination in his subordinates and gives them the maximum authority justified by their understanding of the missions. He establishes an organization which facilitates wholesome personal relationships and which minimizes the amount of coordination required. He motivates maximum effort and directs it into the channels where it is most productive.

It is hoped that this paper will encourage other similar discussions. It may be that these discussions would reveal many psychological approaches which could be applied to practical pursuits of peace time living. It may be that material could be gleaned for psychological courses in business administration and other practical subjects. It may be that many psychologists have been developed into practical administrators and executives capable of outstripping their non-specialized competitors. And, it may be that many occupational pursuits, now considered non-professional, involve so many psychological considerations that they could be more effectively discharged by personnel specialized in the understanding of human behavior.

Received August 20, 1946.

A Statistical Study of Visual Functions and Industrial Safety

N. Frank Stump

General Personnel Department, Revere Copper and Brass Incorporated, Rome, New York

The purpose of this study was to determine the statistical significance of differences between average visual performance for various accident groups. Three groups evaluated from first aid records were compared: (1) An Accident-Free group, (2) A High-Frequency-of-Minor-Accident group, and (3) A Serious-Injury group. The employees were classified in each of these groups by the Safety Engineer, a Personnel Assistant, and the Author, after thoroughly examining the individual accident records for a twelve-month period.

Visual Tests Used

Twelve visual tests were administered to the three employee groups having various past histories regarding accidents. The complete set of tests are incorporated in the Bausch and Lomb *Ortho-Rater.*[1] They are:

1. Acuity, both eyes, far and near distances; 2. Acuity, each eye separately, at far and near distances (not occluded); 3. Vertical and Lateral Phoria,[2] or Muscle Balance, at far and near distances; 4. Color; and 5. Depth or Stereopsis.

Since the tests for "Acuity, each eye separately" were administered without occlusion, both eyes were functioning normally, even though each was being tested separately. The natural-function method is far superior to the obsolete procedure of placing a cardboard before alternate eyes when testing. Unfortunately, this latter method is being generally used today for certifying auto drivers throughout the country, and for testing industrial employees during medical examination.

Results

Table 1 shows Means and Critical Ratios for 22 visual test scores including various score combinations. If raw scores from only twelve Ortho-Rater tests were considered, the number of comparisons between visual functions would have been limited. By combining these raw

---

[1] For description of these tests see: *Standard Practice in the Administration of the Bausch & Lomb Occupational Vision Tests with the Ortho-Rater*, February, 1944.

[2] Joseph Tiffin, *Industrial psychology*. New York: Prentice-Hall, Inc., 1942. Ch. VI.

Table 1

Means and Critical Ratios of Three Accident Groups Compared with Regard to Various Visual Functions

(Note: A = Accident-Free; B = High-Frequency; C = Serious-Injury. N = 108.)

	Means			Critical Ratios		
Tests	A	B	C	M_A-M_C	M_A-M_B	M_B-M_C
1. Acuity—Both Eyes—Far	10.81	10.14	9.83	2.23	1.69	.65
2. Acuity—Right Eye—Far	9.89	8.83	8.86	2.13	2.26	.05[1]
3. Acuity—Left Eye—Far	9.86	8.67	8.61	2.06	2.08	.08
4. Acuity—Better Eye—Far	10.53	9.72	9.69	2.03	2.21	.06
5. Acuity—Worse Eye—Far	9.25	7.78	7.78	2.53	2.50	.00
6. Lateral Phoria—Far	7.44	6.72	7.19	.37	1.25	.67[2]
7. Vertical Phoria—Far	5.56	5.25	4.92	1.75	1.02	.80
8. Acuity—Both Eyes—Near	11.08	10.69	10.58	.08	.87	.21
9. Acuity—Right Eye—Near	9.03	8.58	8.92	.18	.70	.50[3]
10. Acuity—Left Eye—Near	9.01	9.28	9.06	.02	.57	.33
11. Lateral Phoria—Near	6.33	7.14	7.31	1.23[4]	1.02[5]	.21[6]
12. Vertical Phoria—Near	4.61	4.61	4.00	1.84	.00	1.50
13. Lateral Phoria—Near subtracted from Far, plus 10	11.11	9.58	9.89	1.77	1.09	.39[7]
14. Better Eye—Near subtracted from Far, plus 5	5.53	4.64	4.80	1.00	1.63	.50[8]
15. Worse Eye—Near subtracted from Far, plus 6	6.61	5.67	5.58	1.84	1.04	.72
16. Worse Eye—Near	8.64	7.86	8.17	.77	1.24	.44[9]
17. Better Eye—Near	10.00	10.03	9.81	.30	.05[10]	.40
18. Depth	5.64	3.53	4.72	1.31	3.27	1.60[11]
19. Color	4.44	4.25	4.31	.44	.60	.17[12]
20. Vertical Phoria—Near subtracted from Far, plus 3	3.89	3.64	3.92	.08[13]	.90	.78[14]
21. Acuity—Both Eyes—Near subtracted from Far, plus 5	4.72	4.44	4.25	.82	.67	.39
22. Total for 7 tests—Both; Right; Left; Better; Worse; (Far Vision) Lat. Phoria, Near from Far; and Depth	67.08	58.25	59.39	2.09	5.44	.44[15]
23. Education—Grades completed in school	8.53	9.00	8.56	.05[16]	.90[17]	.82

Explanation of footnote numbers: The following combinations should read M_C-M_A 4, 13, 16; M_B-M_A 5, 10, 17; and M_C-M_B 1, 2, 3, 6, 7, 8, 9, 11, 12, 14, 15.

scores into different combinations, however, a total of 22 actual and *alloyed values resulted.*

Educational Status of the Three Groups Compared

The number of school grades completed was taken as an indication of the mental alertness of the individuals. Some studies have been made which indicate that the amount of schooling received is a fair index of general ability. It was regarded, therefore, as essential to hold education fairly constant, on the average, for the three groups. The Critical Ratios between the three injury groups indicate that the average differences in schooling received by these groups are not significant,[3] the Critical Ratios for the Accident-Free group as compared with the High-Frequency and the Serious-Injury group being .90 and .05, respectively. The Critical Ratio of the High-Frequency group when compared to Serious-Injury group was .82.

Accident-Free vs. Serious-Injury

The data reveal that the Ortho-Rater Means show a significant difference in favor of the Accident-Free group over the Serious-Injury group in the following visual functions: (*a*) Acuity—both eyes, far vision; (*b*) Acuity—right eye, far vision; (*c*) Acuity—left eye, far; (*d*) Acuity—better eye, far; and (*e*) Acuity—worse eye, far.

When seven of the visual tests in the Ortho-Rater were combined a highly significant difference resulted between the Mean scores of Accident-Free employees and Serious-Injury scores, the Critical Ratio being 2.69.

Accident-Free vs. High-Frequency

Critical ratios exceeded the criterion when the Accident-Free group was compared with the High-Frequency group on the following test data: (*a*) Acuity—right eye, far; (*b*) Acuity—left eye, far; (*c*) Acuity—better eye, far; (*d*) Acuity—worse eye, far; and (*e*) Depth.

Special mention should be made of the Critical Ratio of 1.99 of Lateral Phoria,—near subtracted from far. Since this Critical Ratio is only .04 below the criterion, a definite trend is unquestionably indicated here.

The combined total of seven visual tests indicates a highly significant difference between the Mean scores of the Accident-Free and of the High-Frequency group, the Critical Ratio being 5.44.

High-Frequency vs. Serious-Injury

Not a single Critical Ratio meets the criterion (2.03) for significant difference between these two groups. These findings are what one should naturally expect, provided the personnel for these experimental groups were properly evaluated. It is generally agreed among authorities in the field that a High-Frequency case is a potentially Serious-Accident case; so, the *present* Serious-Accident individual is, in many cases, the *previous* High-Frequency individual. This thesis is important in the prediction of the eventual Serious-Injury individual on the basis of High-Frequency. This fact is further substantiated by the low Critical Ratio, when a composite of seven visual scores was studied between the High-Frequency and Serious-Injury group.

---

[3] J. P. Guilford, *Psychometric methods*. New York: McGraw-Hill Book Company, Inc., 1936, 548–549. A value of 2.03 is the criterion selected for significance.

Of course, there are some exceptions to the rule. There may be some persons who have serious accidents but who never have been classified as the High-Frequency type. Nevertheless, these data support the above thesis quite well.

Visual functions seem to be an important contributing factor for the safety of the industrial worker. Obviously, many factors enter to make an employee an absolutely safe worker, but visual functions have only recently taken their place as being equally as important as, if not more important than, other factors covered in the literature on safety.

Summary

(1) The Accident-Free group is significantly superior in visual functions to the Serious-Injury group and also to the High-Frequency group in the following: (*a*) Acuity—right eye—far vision; (*b*) Acuity—left eye—far vision; (*c*) Acuity—better eye—far; and (*d*) Acuity—worse eye—far.

This points to conclusions similar to those reached in previous experiments reported by the writer.[4] It is well to note that the Accident-Free group earned significantly higher scores in visual performance for the poorer eye than did the other two groups.

(2) Acuity, Both Eyes, Far Vision scores indicate the Accident-Free group to be superior to the Serious-Accident group.

(3) Depth perception is present to a significantly higher degree in the Accident-Free group than in the High-Frequency group.

(4) In addition, there seems to be a trend in the direction of the visual functions and safety; Lateral Phoria—Near subtracted from Lateral Phoria—Far.

(5) When a composite of seven visual functions was made the Accident-Free group was superior to the High-Frequency and Serious-Injury group, the Critical Ratios being 2.69 and 5.44, respectively.

Received December 7, 1944.

---

[4] Visual functions as related to accident-proneness. *Personnel*, 1944, 21, 50-56; Spotting the accident-prone workers by visual tests. *Fact. Mgmt.*, 1945, 103, 109-112; Visual functions and safety. *Nat. Safety News*, 1944, 49, 22 ff.

Effect of Visual Adaptation Upon Intensity of Illumination Preferred for Reading With Direct Lighting *

Miles A. Tinker

University of Minnesota

A reader rarely hesitates to express a preference for what he considers to be an adequate light intensity for easy and comfortable reading. Luckiesh and Moss (4) claim that this subjective method of investigation yields both significant and interesting results. Although they admit that such results are not desirable for laying a scientific foundation for lighting, it is employed by them as corroborative evidence. In general, preferences for illumination intensities have been employed for two purposes: (1) as supplementary data in prescribing foot-candle standards for reading and (2) to persuade consumers that a high level of light intensity is desirable. The validity of such practices may be questioned. In this paper the plan is to examine critically the evidence in the field and to present new experimental data on the validity of preferred intensities as a measure of ease of seeing.

Review of Previous Experiments

In a study by Luckiesh and Moss (4), 82 subjects selected the intensity of illumination considered as ideal for an extended period of reading black print on white paper. The readers employed their own criteria of comfort afforded by the lighting. Intensities from 10 to 1,000 foot-candles were available. Preferences were distributed as follows: 11 per cent at 10 foot-candles, 18 at 20, 32 at 50, 20 at 100, 17 at 200, 1 at 500 and 1 at 1,000. Although it is stated that the subjects chose an average of about 100 foot-candles, the authors do not note that the median was at 50 in this atypical distribution. It is common statistical practice, of course, to use the median rather than the mean in atypical distributions. In another report (1) Luckiesh states that many observers chose 370 foot-candles of illumination for comfortable reading. The readers viewed an illuminated page of a telephone directory through a small window in a black box. Nothing is said concerning the visual adaptation of the subjects. Little confidence can be placed in this uncontrolled experiment.

Several other experiments have been performed. Luckiesh, Taylor

---

* Grateful acknowledgment is given to the Graduate School, University of Minnesota, for research grant to finance this study.

and Sinden (2), employing well distributed general illumination, found that subjects preferred on the average 5.3 foot-candles for reading 11 and 12 point type, and 10.6 to 16.1 foot-candles for 9 point. For a low degree of contrast between 9 point type and printing surface, 17.4 foot-candles were chosen when 30 were available. In another study with well diffused general illumination, Luckiesh and Taylor (3) obtained preferences of illumination intensities for reading 9 point type. With up to 8 foot-candles available, the readers chose 4.2; with 30, 10.6; with 45, 16.1; with 65, 23.2; and with 100, 35.8. It was concluded that the readers did not choose the maximum because they probably assumed that the desired intensity was somewhere between the extremes. This conclusion appears unwarranted since in the experiment cited above (4), in which preferences were made solely upon the basis of the reader's own criteria of comfort afforded by the lighting and with no suggestion that one's choice should avoid extremes, the median choice was only 50 foot-candles.

Why do such different choices appear when the reader expresses his preference for the illumination intensity best suited for comfortable reading? In the data cited above, there is a suggestion that visual adaptation at the time of the choice plays a dominant role in the choice. As the eyes are exposed to a greater range of intensities, partial adaptation to brighter illumination may occur and this can lead to a choice of more foot-candles for comfortable reading. It is well known that vision readily adapts to easy and efficient seeing within a wide range of illumination intensities. To test the hypothesis that visual adaptation influences the reader's preference for the illumination intensity best for comfortable reading, an experiment was carried out by Tinker (6). There were 144 readers tested. At one sitting a subject was adapted for 15 minutes to 8 foot-candles intensity and then chose, by the paired comparison method, between 8 foot-candles and each of the following intensities: 1, 2, 3, 5, 12, 18, 26, and 41 foot-candles. There was a four minute readaptation to the 8 foot-candles before each succeeding choice. At the other sitting, the subject was adapted to 52 foot-candles and then chose between 52 and each of the following intensities: 18, 30, 41, 46, 59, 62, 71, and 100 foot-candles. When adapted to 8 foot-candles, 8 foot-candles was preferred most frequently for easy and comfortable reading, but the median was at 12 foot-candles. Similarly, when adapted to 52 foot-candles, the median choice was for the intensity to which the subject was adapted. It seems clear that visual adaptation at the moment determines to a large degree the intensity preferred for reading under general illumination.

In a preliminary study of illumination intensities preferred for reading under direct lighting, Aasen, Burwell and Harmon (cited by Tinker, 7) obtained choices in a situation which approximated the light from the old-fashioned type of bridge lamp. The reader explored over a range of from 1 to 50 foot-candles. When progressing from low to high intensities, the mean choice was 21.5 foot-candles; from high to low intensities, 28.9 foot-candles. In a similar study of preferences of illumination intensities for reading, Tinker (7) employed metal shade, flexible arm reading lamps which yielded strictly local illumination. The mean intensity chosen by the readers was 42.6 foot-candles when 3 to 160 were available. Although visual adaptation was not adequately controlled in either of these experiments, there are indications that variation in adaptation was influencing the results. There is need for further experimentation in this kind of situation with strict control of visual adaptation. This is done in the following study.

Description of Experiment

The purpose of this experiment is to determine illumination intensities preferred for easy and comfortable reading under strictly localized lighting when visual adaptation is controlled. The experiment was conducted in a light laboratory with no outside windows, hence rigid control of illumination was possible. Six standard, metal shade, flexible arm reading lamps equipped with regular frosted bulbs provided highly localized lighting. The fairly uniform spot of illumination on the reading stand was about 10 inches in diameter. From the perimeter of this area outward, the illumination decreased rapidly in intensity.

Sixty university students served as subjects. There were two sessions of 50 minutes for each subject. At each session, the subject was adapted to a certain standard level of illumination intensity. At one sitting this level was 20 foot-candles; at the other, 50 foot-candles. Half of the subjects began with the 20 foot-candles, the other half with 50 foot-candles.

The subjects observed one at a time. When a subject arrived at the laboratory, he spent 15 minutes adapting to the standard illumination (20 or 50 foot-candles). Near the end of this period, he was directed to do some sample reading of the printed material that was to be used in the experiment. This text (Winkler's Morgan the Magnificent) was printed in 11 point type with 2 point leading and a 23 pica line width on mat white paper. At the end of the 15 minute adaptation period the following directions were read to the subject:

"I am going to ask you to decide which of several brightnesses of light you prefer for reading. Decide in terms of the light which seems most comfortable for reading. You will be asked to read a few lines under a standard light—the one now on—and then under a comparison light. After trying this for a couple of times, you will choose which light you prefer for reading. Several lights will be compared with the standard."

Ten seconds after notifying the subject that the light now on was one of the lights to be compared, the standard was switched off and simultaneously the comparison light was switched on. Ten seconds later the whole procedure was repeated and then a choice asked for by asking, "Do you prefer this or this?". The subject was readapted for 2 minutes to the standard intensity before the next choice. The 20 foot-candle standard was compared with 4, 10, 50, 75 and 100 foot-candles in a random order. For every alternate subject the random order was different. The 50 foot-candle standard was compared with 4, 10, 20, 75 and 100 foot-candles with a similar randomized presentation. Thus each subject made five choices under each of the two conditions of adaptation.

The illumination intensities were checked every day with a recently calibrated General Electric Foot-Candle Meter to be sure that the intensities were accurate as planned.

Results

The frequency with which each intensity was chosen as best for easy and comfortable reading was computed. Since the standard intensity was compared with each of the five other intensities, it appeared five times as often as each of the comparison lights. It was necessary, therefore, to divide the frequency of choice for the 20 and for the 50 foot-candles by five to make these figures comparable with the others. The results, showing these adjusted data, are given in Table 1. In the first row of each part of the table are listed the intensities of illumination compared. Just below are given the frequencies with which each intensity was chosen as best for easy and comfortable reading of the 11 point type. Percentage frequencies are listed in line three. In reading the

Table 1

Illumination Intensities Preferred for Reading With Direct Lighting

Note: In top section of table, the 20 foot-candle light was compared separately with 4, 10, 50, 75 and 100 foot-candles. $N = 60$ university students. Under these conditions, 4 foot-candles, were chosen 13 times (7 per cent) out of 184 effective choices; 10 foot-candles, 13, etc. The bottom section of the table is read in a similar manner. Here the standard was 50 foot-candles.

	Adapted to 20 Foot-candles					
Foot-candles Compared	4	10	20	50	75	100
Choice frequency	13	13	29	46	42	41
Percentage frequency	7	7	16	25	23	22
	Adapted to 50 Foot-candles					
Foot-candles Compared	4	10	20	50	75	100
Choice frequency	13	13	16	37	35	37
Percentage frequency	9	9	11	24	23	24

table it should be remembered that each intensity was compared with the one to which the subject was adapted.

Examination of the data reveals that visual adaptation at the moment appears to determine to some degree the choices made, but that it does not play the dominant role found by Tinker (6) in the study employing general illumination. For instance, when adapted to 20 foot-candles, there were 29 choices (16 per cent) at 20 foot-candles. But when adapted to 50 foot-candles, there were only 16 choices (11 per cent) at the 20 foot-candles level. Before adjusting the data by dividing by five, there were 145 choices at 20 foot-candles when adapted to that intensity, and 186 choices at 50 when adapted to the 50.

When adapted to either 20 or 50 foot-candles, choices ranged from 4 to 100 foot-candles. But the dominant tendency in both sets of data is to prefer relatively bright illumination in this strictly local lighting situation. Thus, 50, 75, 100 foot-candles were chosen most frequently irrespective of the intensity to which the subject was adapted. The tendency to choose relatively high intensities for reading under strictly localized light agrees with an earlier finding (7).

This preference for relatively bright illumination for reading under strictly localized direct lighting has unfortunate implications for the hygiene of vision. It has been demonstrated that local direct lighting such as employed in this experiment is unhygienic (5). A small spot of bright light surrounded by dimly illuminated areas and shadows in the visual field provides a situation that rapidly produces visual fatigue. It is the lack of uniform distribution of illumination in the visual field that is bad. In a situation of this kind, the greater the intensity of illumination, the worse it becomes for visual work since the demarkation between the small bright area and the dimly illuminated surroundings becomes more pronounced. In fact, the subjects in the earlier experiment on direct lighting (7) were much disturbed by the uneven distribution of illumination in the visual field, especially when higher intensities were employed. Nevertheless the readers do choose the higher intensities under the direct lighting.

Literature on the subject (5) points out that when local lighting is employed it should be supplemented by general illumination to avoid an undesirable degree of brightness contrast within the visual field. This hygienic principle is unknown to many and frequently ignored by others. It is of some importance, therefore, to point out that the undesirable features of such a situation are accentuated as the light intensity is increased. Apparently, therefore, it is unsafe to follow individual preferences as to what seems to be the best intensity for comfortable reading with strictly local lighting. If such a light is used, relatively low intensities produce less visual fatigue.

Summary and Conclusions

1. A survey of the literature reveals a wide range of intensities of illumination chosen as best for comfortable reading under general lighting. Visual adaptation at time of choice largely determines the intensity preferred.

2. The purpose in this experiment was to determine the effect of visual adaptation upon illumination intensities preferred for reading under direct lighting that is strictly local.

3. The illumination used was derived from standard metal shade, flexible arm reading desk lamps. Range of illumination available was from 4 to 100 foot-candles.

4. Sixty subjects observed at two sittings. At one sitting they were adapted to 20 foot-candles; at the other, to 50 foot-candles. After adaptation, each subject chose between the standard to which he was adapted and each of five other intensities on the basis of the intensity preferred for easy and comfortable reading.

5. Visual adaptation at time of choice influenced the choice only moderately.

6. There was a marked tendency to prefer intensities at and above the brightness to which the subject was adapted. This resulted in frequent choice of high intensities.

7. Since, in direct lighting such as used here, bright intensities make a bad situation worse from the viewpoint of hygienic vision, preferences for illumination intensities for reading yield unsatisfactory data for prescribing lighting for the individual.

Received November 22, 1944.

References

1. Luckiesh, M. The eyes should have it (An interview by C. Brooks). *Good Housekeeping*, November, 1934, 88–89, 201–202.
2. Luckiesh, M., Taylor, A. H., and Sinden, R. H. Data pertaining to visual discrimination and desired illumination intensities. *J. Franklin Institute*, 1921, 192, 757–772.
3. Luckiesh, M., and Taylor, A. H. Illumination intensities chosen for reading. *Trans. Illum. Engng. Soc.*, 1922, 17, 269–272.
4. Luckiesh, M., and Moss, F. K. *The new science of lighting*. Lighting Research Laboratory, General Electric Co., Cleveland, 1934, pp. 36.
5. Tinker, M. A. Illumination standards for effective and comfortable vision. *J. consult. Psychol.*, 1939, 3, 11–20.
6. Tinker, M. A. Effect of visual adaptation upon intensity of light preferred for reading. *Amer. J. Psychol.*, 1941, 54, 559–563.
7. Tinker, M. A. Illumination intensities preferred for reading with direct lighting. *Amer. J. Optom.* and *Arch. Amer. Acad. Optom.*, 1944, 21, 213–219.

Mechanical Aptitudes of University Women

Lillian G. Portenier

The University of Wyoming

With wartime economy demands emphasizing the already existing educational and cultural needs for greater mechanical training for university women the Detroit Mechanical Aptitudes Examination was administered to four hundred twenty-five women students at the University of Wyoming during the winter quarter 1942–1943. The revised edition was used. The Detroit test was chosen since, according to the authors, many of the items appeal to girls making the test suitable for both sexes. The same set of norms is used for the two groups. The variation of only one to two per cent between the sexes was not believed to be sufficient to warrant complete sets of separate norms. No appreciable difference could be detected in the median scores.

The similarity in achievement for both sexes for the Detroit test differs markedly from the results found for the Stenquist Assembling Test. Stenquist (8) reports that elementary school girls obtained only about 65 per cent as many correct items as boys, and that women in graduate school obtained about 80 per cent as many correct items as unselected adult men. Similar differences were found in an extensive investigation in which the Minnesota Mechanical Aptitude Tests (7) were used with seventh grade boys and girls, and sophomore men and women in the college of Science, Literature and Arts at the University of Minnesota. The scores made by the university women were about the same as those made by seventh and eighth grade boys. The differences were most marked in a mechanical assembly test and in comprehensive tests of mechanical information. In general, the difference in males over the females increased somewhat in the older groups.

At the present time the available data appear to be inadequate to indicate whether differences in mechanical aptitudes between sexes or between individuals within each sex are due primarily to differences in general ability or to a special ability or abilities. Freeman (5) states, "One may well question whether it is appropriate to speak of mechanical ability as dissociated from general ability, that is, as a special ability." The results of the Detroit Mechanical Aptitudes Examination administered to the women students at the University of Wyoming were analyzed largely for the purpose of throwing further light on this question.

Of the women tested 48 per cent were freshmen, 24 per cent sophomores, 15 per cent juniors, and 13 per cent seniors. There are eight subtests in the Detroit Examination. Tests two and eight were designed to measure Motor Ability; tests three, five, and seven, Visual Imagery; tests one and six, Mechanical Information, and test four, Arithmetic. The "norms are based on ten thousand scores, chiefly the unselected groups of pupils at the eighth and ninth grade levels in Detroit. However, the examination has been given also to several hundred mentally subnormal children, to some normal fifth and sixth grade pupils, and to small groups in the senior high school as high as the twelfth grade."

Table I

Medians, Means, Standard Deviations, Ability Ages and Ratings for the Scores on the Detroit Mechanical Aptitudes Examination, and the Ohio State University Test, for University Women

Test	Number	Median	Mean	S.D.	Age	Ratings
Detroit Mechanical Aptitudes						
Part I, Motor Ability (Tests 2, 8)	425	60.00	59.80	8.45	19-3	A, Very Superior
Part II, Visual Imagery (Tests 3, 5, 7)	425	91.11	85.25	17.55	17-2	B, Superior
Part III, Mechanical Information (Tests 1, 6)	425	50.29	49.85	8.85	16-9	B, Very Good
Part IV, Arithmetic (Test 4)	425	30.87	30.94	6.30	18-8	B, Very Superior
Total Scores (Tests 1–8 inc.)	425	229.26	225.68	30.60	17–5	B+, Superior
O.S.U. Psychological						
Part II, Analogies	396	30.56	31.00	11.65		53%ile
Total Scores (Parts I–II–III)	396	80.00	80.25	25.31		55%ile

Since the Ohio State University Psychological Test is administered to all students at Wyoming as high school seniors or after enrolling at the University the results of this test on file in the personnel department were used as a basis for determining the general mental ability of the women used in this study. The medians, means, standard deviations, mechanical ability ages and ratings for the scores on the subtests and on the total test for the Detroit Examination, and for Part II, analogies, and for the total score on O. S. U. Test are shown in Table I. The mean percentile rank on the O. S. U. Test for the group is strong average on the basis of the national norms. In comparison with the norms for the Detroit Examination the median and mean scores were above average on all of the subtests and also for the total score on the test. The lowest mean score was achieved on the tests which were designed to measure Mechanical Information and the highest score on the tests for Motor Ability. The mean score for Part IV, Arithmetic, also was very superior. As indicated by Table 2, 72.5 per cent of the women earned superior scores on the mechanical aptitudes tests, and with the exception of but one woman the remaining 27.5 per cent earned average scores. This group of women was above average in mental ability as well as in mechanical aptitudes and a definite relationship between the scores was evident.

The superior results for this group of women may be due in part to the fact that the norms of the Detroit Examination are based chiefly on unselected groups of junior high school pupils. However, in no case did the test fail to measure the limit of ability on any of the subtests since none of the women earned perfect scores on any of them.

Table 2

Rating of 425 University Women on the Detroit Mechanical Aptitudes Examination and the O.S.U. Test

Rating on O.S.U. Test	Rating on Detroit Mechanical Aptitude Test		
	Inferior Below 20%ile	Average 20 to 80%ile	Superior 80%ile and Above
Superior 80%ile and Above	0%	.7%	15%
Average 20 to 80%ile	.2%	20%	55%
Inferior Below 20%ile	0%	6.6%	2.5%

To study further the relationship between the results on the tests for mechanical aptitudes and mental ability, product-moment correlations were computed between the scores on various tests and subtests. These correlations are shown in Table 3. With the exception of the tests for mechanical information and motor ability the correlations found are higher than those found for most other similar studies in which comparisons were made between mechanical aptitudes and general mental ability.

Stenquist (8) reports a correlation of .23 between the Stenquist Assembling Tests and a composite score from six verbal intelligence tests in a group of 267 seventh and eighth grade boys. The same correlation viz. .23 was found in this study between the subtests for Motor Ability and the total O. S. U. Test. The Stenquist Assembling Tests are far more mechanical in character than the Detroit tests. They are not of the paper and pencil type. They involve the construction of common mechanical objects from given parts such as a mouse trap, door lock, clothespin, and bicycle bell. Since a correlation of .23 indicates a very slight degree of relationship Stenquist concludes that mechanical aptitude seems to be a special ability. A correlation of .13 was found by Paterson (7) for 100 junior high school boys between I. Q.'s estimated from socres on the Otis Intelligence Test and the scores obtained from a mechanical aptitude battery, composed exclusively of apparatus or manipulation tests. For the Minnesota Paper Form Board Test, which is a paper and pencil test of the ability to handle spatial relations, and a vocabulary test Anastasi (1) obtained a correlation of .07 for 225 male college students.

The O. S. U. Test and the Detroit Examination appear to have more in common than the tests in the studies just cited. The higher correlations found may be attributed to the common influence of the comprehension of verbal directions, knowledge of words, and general facility with verbal material. This conclusion seems to be supported by the relatively high correlations found between certain subtests and the total of the Detroit Mechanical Aptitudes Examination and the total scores and Part II, Analogies, of the O. S. U. Tests. As shown in Table 3 the correlation of the total O. S. U. Test with Visual Imagery is .42, with Arithmetic .43 and with total Mechanical Aptitude scores .53. Between Visual Imagery and Analogies it is .38, Arithmetic and Analogies .40 and between the total Detroit test and Analogies it is .47. The authors report a correlation of .65 between the Mechanical Aptitudes Examination and the Detroit Advanced Intelligence Test.

Table 3

Correlations Between Scores on the Ohio State University Test and the Detroit Mechanical Aptitudes Examination for University Women

Detroit Mechanical Aptitudes Examination	Total O. S. U. Test	Part II, O. S. U. Analogies
Part I, Motor Ability	.23 ± .032	
Part II, Visual Imagery	.42 ± .026	.38 ± .028
Part III, Mechanical Information	.09 ± .033	
Part IV, Arithmetic	.43 ± .025	.40 ± .027
Total Test, Mechanical Aptitudes	.63 ± .024	.47 ± .024

Part I of the O. S. U. Test is largely a measure of vocabulary; Part II, analogies, was designed to measure a combination of reasoning ability, spelling, capacity for logical associations and knowledge of grammatical forms; and Part III is primarily a test of ability to read difficult passages. Part II would seem to have most in common with the Detroit Examination since little vocabulary is involved and none of the subtests call for ability to read difficult passages.

The findings from the data presented in the present study tend to support Guilford's (6) conclusion. He states, "Strange as it may appear, there seem to be no special abilities of speed, learning, musical talent or mechanical bent. ———Musical talent is very complex and so is mechanical skill." Thus, while the correlations between verbal and numerical tests generally have been found to be low a higher relationship may result when both tests involve similar abilities, such as ability to follow directions as appears to be the case, to some extent, with certain subtests in the Detroit and Ohio tests, and to an even greater extent between the Detroit Advanced Intelligence Test and the Mechanical Aptitudes Examination.

That mechanical aptitudes are a complex of skills and abilities rather than a unitary special ability is further indicated in a study of "Sex differences in the understanding of mechanical problems" by Bennett and Cruikshank (4). They found women slightly superior to men in tests of manual performance if dexterity is involved, and men clearly superior in tests of mechanical assembly, or if strength is needed. Again, the superiority of men was not so marked in tests involving perception of spatial relations as the Minnesota Paper Form Board or the Minnesota Spatial Relations Test. Subtests 2 and 8, designated as tests of motor ability in the Detroit Examination, involve dexterity, and subtests 3, 5 and 7 termed tests of visual imagery depend to some extent on the perception of spatial relations. Bennett and Cruikshank (3) state, "Tests which require speed and accuracy in simple perceptual discriminations and in simple motor tasks show slightly better scores for girls. In tests of strength, boys are superior. In more complicated perceptual and spatial tasks, boys also score higher. ———To a large extent there can be no doubt that the superiority of boys in tests of mechanical ability is a function of our present system where women are ordinarily not expected to deal with certain types of mechanical contrivances nor encouraged to do so. When it happens, as it does in a war period, that women are called upon to do some of these things, their performances are far more creditable than might have been expected."

Just as an I.Q. derived from one battery of mental tests does not signify the same thing necessarily as an I.Q. from another battery so the scores resulting from different tests of mechanical ability may be indicative of widely varying aptitudes or skills. The results analyzed in this study seem to indicate that the mechanical abilities measured by the Detroit Mechanical Aptitudes Examination are not dissociated completely from general college ability as measured by the O. S. U. Test when the abilities measured are those of university women.

Received December 1, 1944.

References

1. Anastasi, A. A group factor in immediate memory. *Arch. Psychol.*, 1930, 120, 1–61.
2. Baker, H. J., et al. *Manual of directions, Detroit Mechanical Aptitudes Examination, Form A.* Bloomington, Ill.; Public School Publishing Co., 1940. Pp. 16.
3. Bennett, G. K., and Cruikshank, R. M. *A summary of manual and mechanical ability tests.* The Psychological Corporation, 1942. Pp. 80.
4. Bennett, G. K., and Cruikshank, R. M. Sex differences in the understanding of mechanical problems. *J. appl. Psychol.*, 1942, 26, 121-127.
5. Freeman, F. N. *Individual differences.* New York: Henry Holt and Co., 1942, pp. 300–301.
6. Guilford, J. P. *General psychology.* New York: Van Nostrand Co., 1942, pp. 515.
7. Paterson, D. G., et al. *Minnesota mechanical ability tests.* Minneapolis: University Minn. Press, 1930, pp. 586.
8. Stenquist, J. L. *Measurements of mechanical ability.* N. Y.: Teachers College, Columbia Univ., 1923, pp. 101.

Aptitude and Interest Patterns of Art Majors in a Liberal Arts College

Dorothy M. Barrett

Hunter College of the City of New York

Many students in college have difficulty in choosing a field of specialization. Often young people who succeed in electing a field of specialization later change their minds one or even several times. Even when they make decisions and follow through with them, students frequently later regard their choices as mistakes. Much energy and ability are wasted because students fail to find fields of study and work for which their aptitudes and interests suit them.

This paper describes a study, part of a larger project, which was undertaken with the intention of trying to discover whether or not a group of juniors and seniors who were majoring in art in a liberal arts college could be differentiated by means of a battery of tests from a group of juniors and seniors majoring in other fields within the same institution. The study was undertaken with the assumption that a *battery* of tests would be necessary to obtain satisfactory predictions inasmuch as success in an academic or occupational field is likely to depend upon a number of aptitudes and interests and not upon any single ability.

If it proved possible to differentiate art majors from a group of control subjects by means of psychological tests, then it was proposed, after further checking the results, to use the battery of tests as diagnostic measures in advising underclassmen who consider art as a major but who, in spite of having been accepted by the art department, still feel uncertain about the wisdom of their choice.

Psychologists who are concerned with the problems of the vocational and educational guidance of students in colleges of liberal arts may find something helpful in the results of this project. The critical scores obtained in this study may not be the best ones for another institution because the predictive value of any scores would differ for each institution, depending upon the curriculum in which the test program is used, the number and types of specializations offered, the admission and academic standards of the college and the kind of art training given. However, a report of the results on those tests which proved useful in identifying art majors may prove valuable to other workers faced with advisory problems.

Subjects

In response to a request for volunteers, forty students in the art major and an equal number of upperclassmen specializing in other fields took the battery of tests. The art majors were students who, as juniors and seniors, had obtained satisfactory grades in their field of specialization and were quite satisfied with their choice of art as a major. The control group represented juniors and seniors in other majors who also expressed satisfaction with their fields of specialization. Subjects in both groups were interested in the tests, having been advised that they could learn their scores at a future date. All students were girls, students in Hunter College of the City of New York.

Tests [1]

The tests which were administered to both the art majors and the control subjects included Meier's Test of Art Judgment (1940 revision), Strong's Vocational Interest Blank for Women, the Study of Values, Series BB of the Revised Minnesota Paper Form Board and the Guilfords' Prognostic Test for Students in Design.[2]

Scores were available for all subjects for the Psychological Examination of the American Council on Education, administered at the time of admission to college. For a considerable number in the group, records were available of the scores on Greene's Michigan Vocabulary Profile Test which had been administered shortly after a student's admission.

Results on Meier Art Judgment Test

The distribution of test scores made by the art majors on the Meier Art Judgment Test overlaps markedly with the distribution of scores for the control subjects. As a group, however, the students specializing in art make higher scores than do the control subjects. The average score for the art majors is 109 while the average for the control group is 103. Using Fisher's t-test for measuring the significance of the difference in the means of independent small samples, t equals 2.6, indicating that the difference is significant at the 1% level.

There are other more useful differences between the two groups, however. A study of the distribution of ungrouped scores resulted in the establishment of two critical scores. A minimum score of 99 eliminates a fair number of control subjects while discriminating against only two art majors. A score of 107 and over singles out a considerable proportion of the art majors while including only a small number of the control group. Thus, there are three significant ranges of scores.

---

[1] The author is indebted to Miss Adele Linker for assistance in scoring the tests and to the Bureau of Educational and Vocational Guidance of Hunter College of the City of New York for sponsoring this study.

[2] Guilford, J. P., and Guilford, R. B. A prognostic test for students in design. *J. appl. Psychol.*, 1931, 15, 335–345.

Each subject in the experiment was classified in one of the three ranges of scores, according to whether her score was below 99, from 99 to 106 inclusive, or 107 and over. Figure 1 shows the results of such a classification. The number of control subjects in each range is indicated by the figures to the left of the line within the bar. The number of art majors in the same range of scores is represented by the figures to the right of the line within the bar. The per cent at the left end of each bar is the percentage which the control group represents of the total group scoring within the limits indicated. The per cent at the right of each bar is the percentage which the art group represents of the total group scoring in that category.

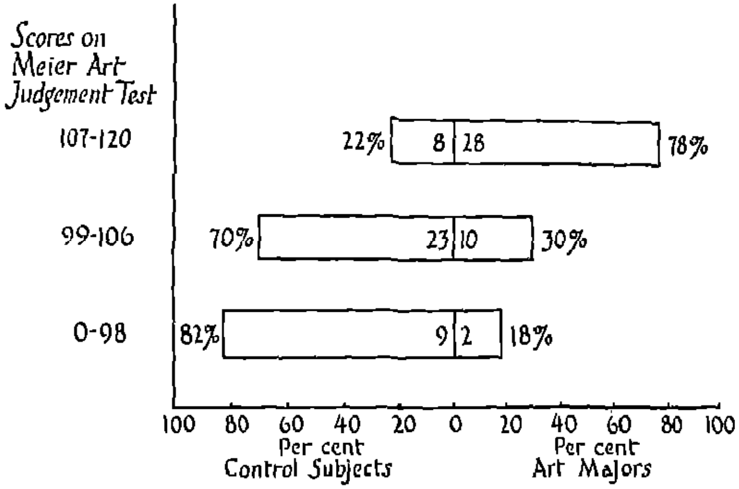


FIG. 1. Showing how many were control subjects and how many were art majors of those students scoring within certain ranges of scores on the Meier Art Judgment Test.

Reference to Figure 1 indicates that 11 subjects had scores below 99 on the Meier Art Judgment Test. Of this group having low scores, 82% are control subjects and 18% are art majors. Of the subjects having scores from 99 to 106 inclusive, 70% are control subjects and 30% are art majors. On the other hand, of the subjects with scores of 107 and over, only 22% are control subjects while 78% are art majors.

On the basis of this study, any student scoring up to and including 98 on the Meier Art Judgment Test would have only 18 chances in 100 of being a successful and satisfied art major in Hunter College. If her score is between 99 and 107, her chances of being an art major in the same institution are still only 30 in 100. However, if a student has a score of 107 and over, her chances of being an art major are more than double, then being 78 in 100.

Results on Vocational Interest Blank for Women

The usual method to be followed in attempting to differentiate between the two groups in this study would be to use only the ratings for Artist on Strong's Vocational Interest Blank for Women, with the expectation that there would be more high ratings for Artist among the art majors than among the subjects of the control group. However, the author believed that high scores on occupations other than Artist might show either a positive or possibly a negative relationship to success in art. In other words, a student with an average rating as an Artist and no higher ratings for other occupations might succeed in art, while a student with an average rating as an Artist but a higher rating in one or more other occupations might not be a person who would succeed in art.

Table 1

Showing the Distribution of Ratings for Artist on Strong's Vocational Interest Blank for Women

	Ratings for Artist					
	C	C+	B−	B	B+	A
Art Majors	2	2	2	7	9	18
Control Subjects	6	9	7	7	5	6

Because, then, there was a possibility of finding certain additional relationships which might differentiate between the two groups, the blanks were machine scored for all occupations. The writer studied carefully the distributions of letter grades for the art majors and for the control subjects for each of the 17 occupations on the test. The ratings for several of the occupations did differentiate between the two groups. It was concluded, however, that the additional differentiation on the basis of the extra scores was insufficient to warrant, for future counseling services, the expense of the extra scoring required. The remaining discussion, therefore, is limited to a consideration of the scores for Artist.

The art majors and the control subjects obtained rather different ratings for Artist. The tabulation of the letter ratings for each group is given in Table 1. Reference to the table indicates that only six of the forty art majors scored B− or less as Artist while twenty-two of the control subjects had such low ratings. Thirty-four of the art majors scored B or better, while only eighteen of the control subjects had such high ratings. In other words, high scores for Artist on the Strong test are more often than not associated with successful specialization in art.

Figure 2 shows that a student's chances, according to the results of this study, of being a successful and satisfied art major are only 21 in 100 if she has a rating of B− or less. If she rates B, B+ or A, her chances of being a satisfactory and satisfied art major become 65 in 100.

Results on a Study of Values

Keeping the control and experimental groups separate, a study was made of the distributions of the scores for each value on the Allport and Vernon Test.

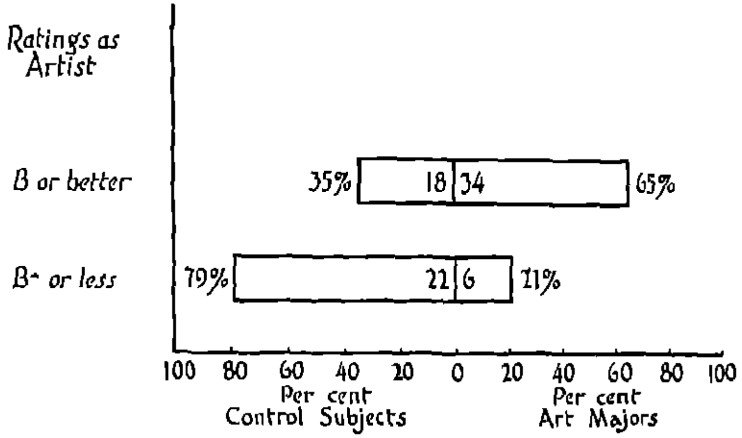


FIG. 2. Showing how many were control subjects and how many were art majors of those students scoring within certain ranges of ratings for Artist on the Vocational Interest Blank for Women.

The art majors and control subjects, while yielding overlapping distributions, did nevertheless show certain differences. Both groups had very high scores on the Aesthetic scale, but a greater number of the art majors scored at the extreme end of the scale than did the control subjects. More control subjects had extremely low scores on the Aesthetic value than did the art majors. Taking into consideration the relative number of art majors and control subjects at each end of the scale, a score of 46 and over was given a value of +2 and a score of 32 or less was assigned a weight of −3. In like fashion, weights were worked out for the other values on the test. On the basis of these weights, it was theoretically possible to have a total weighted score for the Study of Values varying from −17 to +12. Actually, the scores varied between −13 and +8.

A study of the distributions of the weighted scores led to the establishment of certain critical scores which were used to set up ranges of scores that were significant. The extent to which the experimental and control groups differed is indicated in Figure 3.

A total of 25 students had scores between −13 and −2. Of this group, 88% were control subjects and only 12% were art majors. On the other hand, in the group having very high scores of 4 and over, 86% were art majors and only 14% were control subjects. The middle range of scores included more equal proportions of art majors and control subjects, the exact proportions being shown in Figure 3. (The figures inside the bar again indicate the number of cases upon which the per cent is based.)

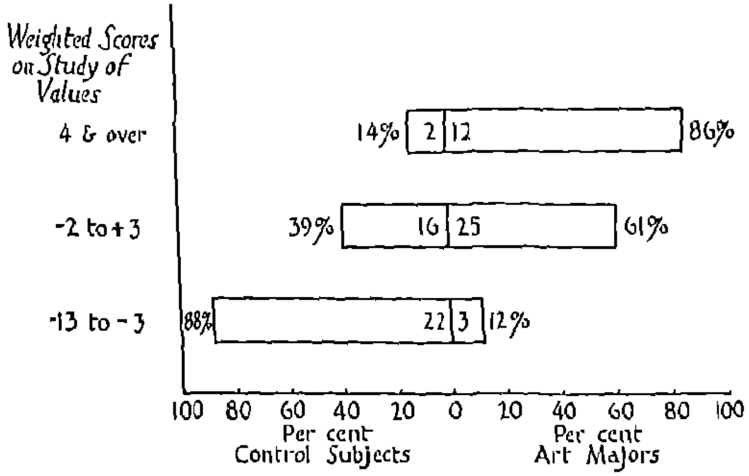


FIG. 3. Showing how many were control subjects and how many were art majors of those students scoring within certain ranges of total weighted scores for the Study of Values.

Weighted scores on the Allport and Vernon Test differentiated well between the art majors and the control subjects. On the basis of this test alone it is possible to give a fair estimate of a student's chances of being an art major. The average score for the art majors was +2 and for the control subjects it was −3. Again using the *t* for small samples, we find that the difference between the averages is significant at the 1% level.

Results on Revised Minnesota Paper Form Board

Series BB of the Revised Minnesota Paper Form Board was included in the battery of tests administered to the subjects on the assumption that the ability measured might be one of the aptitudes contributing to successful composition, as well as contributing to success in such specialized courses in the art major as drafting and lettering.

The control and experimental groups had rather similar scores on this test of spatial relations. The average score for the control subjects was 43 and for the art majors was 47, a difference which is small but significant at the 5% level. As indicated in Figure 4, 79% of the group scoring below 41 were control subjects, while only 21% of the group were art majors. At the other end of the distribution, 72% of those students scoring 51 and over were art majors and only 28% were control subjects. The middle range of scores included approximately half art majors and half control subjects.

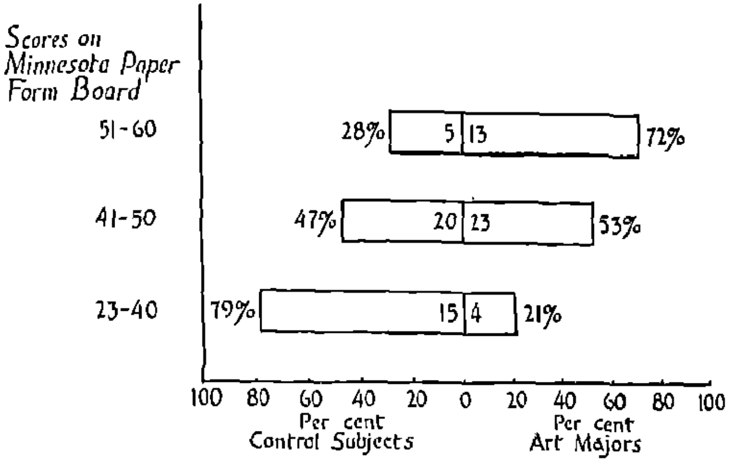


FIG. 4. Showing how many were control subjects and how many were art majors of those students who scored within certain ranges of scores on the Minnesota Paper Form Board.

Extreme scores on the Revised Minnesota Paper Form Board test were significant in differentiating between art majors and control subjects. This test gave results sufficiently diagnostic to be included in a final battery of tests.

Results on Other Tests

The experimental and control groups had identical Q, L, and Total scores on the A. C. E. Psychological Examination. The two groups did differ, however, on the Michigan Vocabulary Profile Test. Analysis of the profile scores *made by the students when they were freshmen* indicates that for the Fine Arts Vocabulary section the average score for the students who later became art majors was higher than the average score of those students who later majored in other fields. Because scores were not available for all subjects, however, the test was not included as one of the predictive measures in the final battery of tests.

The Guilfords' Line Drawing Test, which was included in the experiment, brought out some interesting differences between the two groups. The test appeared to measure a certain creative ability which the other tests did not touch. However, the subjectivity of the scoring of the test proved too great an objection to its use in practical circumstances.

Results of Four Tests Combined

The scores from Meier's Test of Art Judgment, Strong's Vocational Interest Blank for Women, the Study of Values and the Revised Minnesota Paper Form Board were weighted in terms of the extent to which each test differentiated between art majors and control subjects. When the scores were combined, the results were strikingly satisfactory.

The combined weighted scores ranged from -20 to $+18$. Dividing the subjects into two groups on the basis of whether a given score was below zero on the one hand, or zero or above on the other hand, we find that the control subjects are separated from the art majors with 85% accuracy. The actual number of art majors and control subjects within each scoring range is indicated in Table 2.

Table 2

Showing the Distribution of Total Weighted Scores for Test Battery

	Total Weighted Scores -20 to -1	 0 to $+18$
Art Majors	4	30
Control Subjects	32	8

Ordinarily, psychological tests pick out the failures more readily than the successes. A minimum score is often necessary for success. Not as often does a high score mean success. The tests used in this study have, however, tended to identify the art majors in positive fashion as well as to identify those students least likely to possess such a pattern of aptitude and interest.

It should be pointed out that some of the scores made by the art majors may be spuriously high. It is possible that the scores on the test of Art Judgment and on the Revised Minnesota Paper Form Board may be higher than would ordinarily be the case just because the students have been studying art. To what extent the differences established in this study would be substantiated if subjects were given all tests at the beginning of their college courses is at present a matter of conjecture. The author is carrying out a study to check on this point.

Meanwhile, however, the distribution of the combined scores warrants further consideration. According to the results indicated in Figure 5, no art majors were included among the eighteen students having *the lowest scores, the tests thus predicting 100% accurately* for this range of scores. Eighty-five per cent of those subjects having high scores were art majors, a very satisfactory degree of differentiation between the two groups.

A study of the five control subjects included in the range of scores which was generally typical of art majors leads the writer to wonder if the tests may not be even more successful than the figures have indicated. Reference to interview notes jotted down at the time the subjects took the tests or reported for their scores shows that one control subject who scored high on the battery of tests was an art major originally, but changed to another field because of a personal experience which had turned her against the study of art. It seems fair to conclude that she had reacted against personalities and not against art as a major. This girl was evidently not a fair control subject if what we are trying to predict is interest and ability in the field of art.

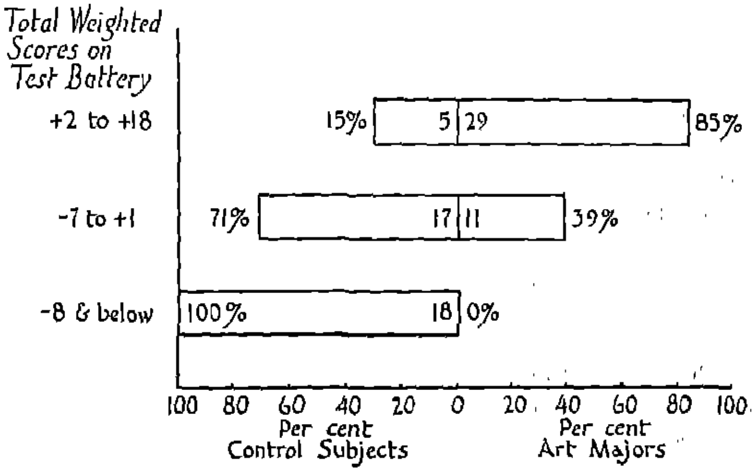


FIG. 5. Showing how many were control subjects and how many were art majors of those students scoring within certain ranges of composite weighted scores on the test battery.

Another control subject just recently discovered a flair for art. She is continuing her present major but has decided to include as many art courses as possible in her program. Very probably she would have made a satisfactory art major if she had become aware of her aptitudes and interests earlier.

A careful study of the aptitude and interest patterns of several of the art students who fell outside of their group on the basis of the total weighted scores on this battery of tests leads the author to conclude that although the students are satisfied with their majors, they might have done better work and been better satisfied in other fields. For example, one art major seems to have a flair for science, a fact which she never realized and never followed up. It is possible that it is significant that this student had scores more similar to the scores of the control subjects than to the scores of the art majors.

In conclusion, then, it can be stated that the total weighted scores on this battery of tests differentiated clearly between the art majors and the control group. The difference between the averages of the two groups for the total weighted score is significant at the 1% level.

Summary

1. A battery of tests which included Meier's Test of Art Judgment, Strong's Vocational Interest Blank for Women, the Study of Values and the Revised Minnesota Paper Form Board differentiated between a group of 40 art majors and a group of 40 control subjects with 85% accuracy. Critical scores established for each test made it possible to combine the results of the four tests and then to predict with only a small margin of error a student's likelihood of being an art major in the curriculum of Hunter College of the City of New York.

2. Other tests studied but not included in the final battery of tests included Greene's Michigan Vocabulary Profile Test, the Guilfords' Prognostic Test for Students in Design and the A. C. E. Psychological Examination.

3. No differences were found between the control and experimental groups for either the Q, L, or Total scores on the Psychological Examination. On the Michigan Vocabulary Profile Test, which had been administered to many of the students at the time of their entrance to college, the art majors ranked higher than the control subjects on the Fine Arts Vocabulary section. Because scores were not available, however, for all 80 subjects, the test was not included in the final battery of tests for which a weighted total was obtained.

4. The Guilfords' Line Drawing Test was administered to all subjects. A kind of creative ability seemed evident in the data, suggesting interesting possibilities for further analysis. However, evaluation of the records for this test is too subjective at this stage to warrant the inclusion of the test in a battery of tests for immediate practical use.

Received November 22, 1944.

Merit Examination Cut-Offs and Weights

Wm. Thos. Toolan

Supervisor of Recruitment, Los Angeles City Board of Education

Although the Merit System has made much progress during the last few years, it still faces the problem of shaking off some of its inheritances such as setting weights and passing marks in advance of examinations, and making cut-offs without careful study. Some public personnel agencies are compelled to arbitrarily set weights ahead of examinations because of the limitations imposed upon them by the laws under which they operate; other agencies accept this sort of procedure because others use it. The procedures set forth in this article will be found to be fairly objective because of the use of simple analysis and study of parts of examinations and their relation to each other.

A "cut-off," for the purpose of this article, is a line on a frequency distribution of scores below which scores are considered as not acceptable or are below a predetermined standard. The lower tail of the distribution is usually cut off. The upper tail represents persons supposedly "too good" for the particular job. This upper tail is not considered when making cut-offs in public service.

The reason for these eliminations is not always understood by the job seeker. Most of them feel that they are qualified and therefore should be on the employment list. The recruitment office, of course, has definite reasons for making an elimination after each part of an examination, and just before the eligible list is established.

Applicants are not always told the real reason why their names do not appear on the list. They are often told that "the size of the list did not warrant placing their names upon it." This would imply that "the size of the list is based on our anticipated needs." No matter how a notification of standing, that tells the applicant his name is not on the list, is worded, it means to him that he has failed in the examination. This generally makes for ill feeling with the public.

Elimination at each step of the examination is ordinarily done for the purpose of getting rid of unfit applicants. The final elimination should be made on the basis of including all qualified applicants. That is, no matter what the needs or anticipated needs might be, all qualified applicants should appear on the eligible list. The excuse of the recruitment or assignment office that it entails some additional clerical work hardly seems sufficient. For the sake of public relations, it is well that elimination of applicants seeking jobs in the public service be done more carefully and scientifically than heretofore.

Before starting to make a cut-off, it would seem that a knowledge of what had taken place to give us these scores in the distribution would be a *sine qua non*. If we have not formulated the specifications, drawn up the test, administered or analyzed it, we should get in touch with those who have done so. We are now qualified to take the first step in making the cut-off. This tentative cut-off will be a guess, backed up by the supporting data which we have at hand.

A Tentative Cut-Off

Around, or somewhat above and below this tentative cut-off, will be found test papers none of which are exactly alike. If the test is a written one or a practical (performance), the results for each candidate will be different although they may have the same score. One will have missed items which the other or others will have gotten right. If we have made too low a cut-off, we will be in the midst of a lot of poor papers. If we have made too high a cut-off, the papers will, generally speaking, be quite acceptable. Thus, we may have to move up or down and make several test cut-offs before we arrive at a place where we are fairly sure that we have neither done any great injustice nor included any obviously unfit papers.

If there are many vacancies or anticipated vacancies and there is a dearth of applicants, we may be forced to include more above the cut-off than we would otherwise. Or, if there are few vacancies, or anticipated vacancies, we will probably not want to carry an excessive number through the remaining parts of the examination. There is great danger in picking up the unqualified when making a low cut-off, and there is danger in building ill feeling by failing a large group of qualified persons when making an unnecessarily high cut-off.

A Typical Example

What one runs into can be seen by examining a group of practical papers for say stenographers. . Among some obviously unfit papers is found a paper which on close examination shows that the contestant did very poorly at the beginning of the test and gradually got better toward the end in spite of the fact that there was an increase in words per minute. This was not due to lack of "warm-up" as we find that ample practice dictation was given prior to taking the test. To those who are experienced in giving stenographic tests, these thoughts would immediately come to mind: Taking in shorthand and transcribing are different steps. The longer the test, the greater the possibility of the first part of the notes getting cold. The testee who lacks a vocabulary of sign words or lacks retention will not be able to read much of the first part of her notes because they are colder. No matter how fast one can take shorthand, if she cannot transcribe or read her notes when they are cold, they are of little value to anyone.

Many limitations are imposed on one who is attempting to make a scientific cut-off for a practical test. It is impossible to assume a purely objective attitude. Persons scoring typing or stenographic tests do not rate exactly the same. Testees often make the same error but under different circumstances. One stenographic testee omits five words of dictation and patching it up submits an acceptable letter, another makes the same error but makes no attempt to "fill in." They are penalized alike.

Although approximately the same *modus operandi* may be used in making cut-offs, no two are made in exactly the same manner. The more we study the peculiarities of the papers, whether written or practical, the more we are in doubt regarding the manner in which this or that answer was chosen in an item of a written test, or just why this otherwise good practical test was spoiled by the testee doing something unusual or out of the ordinary.

Ordinarily an examination includes a written and an oral part. Sometimes a practical test follows the written, and occasionally a practical and oral is used. This last combination is used where the practical is broad enough to test for the degree of skill needed for the job. However, where it is necessary to test for technical knowledge, a written test is used. The oral test is ordinarily used as a test of general fitness for the job. It includes evaluation of personal characteristics, and education and experience.

When Parts Should be Weighted

The weighting of parts of an examination before it is held can hardly be anything other than an estimate of the results of the examination. For instance, the written is compiled and intended to test for certain things, but after it is administered, and the results thereof studied, it is found that it has failed to come up to expectations. The distribution of scores may show clearly that it is too easy or too difficult for the group. This would, then, make a difference in the weight assigned it. Ineffective rating on the part of oral or practical test raters should be cause for lowering the weights of these parts, whereas otherwise they would be assigned higher weights.

Where an agency uses the same written and other parts of an examination, and the same rater requirements, and the quality of the applicants remain the same, about the same results can be expected. This being the case tentative weights may be set subject to study after all parts of the examination have been completed.

The weight of an examination part will vary when studied in the light of the requirements for the job. Examples of extremes, say for personal characteristics requirements, would be "secretary" and "billing machine operator." Although studies have shown that the majority of employees are unsatisfactory because of personal characteristics, we must admit that the degree to which personal characteristics are important will vary with the job. Thus weights for the parts of an examination will depend upon the nature of the position for which the examination is given.

The question as to the effectiveness of a part in an examination probably is the least objective step in determining its weight. All considerations of the examination are kept in mind and the paired comparison method of ranking employed. In other words, this phase of the study presupposes a thorough knowledge of all that has any bearing or throws any light on the examinations. It is backed up by applicable information, or supporting evidence.

If the written test has been administered and studied as to reliability, it can be relied upon to give about the same results each time it is used, provided the requirements for admission to the test have not been changed materially. The practical is probably the easiest of all the parts of an examination to size up as to its effectiveness, depending, of course, upon the raters, the nature of the test, how elaborate it is, etc. In all parts of an examination, with the exception of the written, evaluation of effectiveness is based on observation, and, of course, distribution of scores.

As soon as we affix a weight to a part of an examination we indicate that it has been considered in relation to the other part or parts of the examination. Its importance or effectiveness, then, is reflected in the weight assigned it. When an attempt is made to consider a part without first considering the other parts, the other parts may be left with too small or too large weights. That is to say, if we insist that written parts have low weights affixed to them, we are apt to find that the oral interview parts have been given excessive weights.

What Weights Do

If a part of an examination is important or effective, it should be weighted accordingly. When a part has proven itself worthy of the heaviest weight in an examination, and such a weight has been assigned it, it becomes the controlling factor of the relative placement of persons in the final distribution, or the eligible list. Thus the greatest weight in an examination will tend to move persons on the final distribution a greater number of places than will the smallest weight assigned. What weights will do can be brought home most effectively by reversing assigned weights and observing the results.

Frequency distributions of parts of examinations vary to the extent that no two are identical. This means, then, that their spreads and "forms" are instrumental in shaping final distributions and eligible lists. A distribution of a part which has a tendency to be bunched (skewed) upward or downward or for any reason does not have a good spread of scores has set a comparatively low weight for its part, and should be weighted accordingly irrespective of what weight has previously been forecast for it.

The setting of weights for examinations in other than a careful manner is apt to spoil everything that has gone before. Such *a priori* methods as are sometimes used because of carelessness or ignorance of a few simple rules, are not in accordance with the merit system because they have a tendency to place persons on eligible lists out of the order of their relative excellence. This is true of any procedure which is not done in the manner in which it should be done. The results not being apparent to the layman he does not challenge the methods employed.

Received October 12, 1944.

Book Reviews

Rosenstein, J. L. *The Scientific selection of salesmen.* New York: McGraw-Hill, 1944. Pp. xii + 259. $3.00.

The audience for, and major objective of, this book is stated as follows by the author: "This book is for sales executives who are long-range minded. Its purpose is to describe, to explain, and to teach the steps in a scientific salesman selection program" (ix). For many reasons the publication fails by a wide margin to meet either the needs of its intended audience or its primary objectives.

Throughout the book there are evidences of hasty, superficial writing. Important concepts are ignored or treated so sketchily that they might better have been ignored. On the other hand, certain simple ideas, such as the computation of an average, are heavily labored. Part of the difficulty encountered by the author arises from his attempt to oversimplify psychological tools and techniques which just cannot be so treated for lay readers. An example of this is the attempt to teach interpretation of complex psychographs to laymen in fewer than forty pages (211–244). In the section on interviewing, such terms as "substitution," "persecution," and "projection" are introduced with little explanation for naive readers.

Treatment of tests and testing is entirely inadequate. Two quotations suffice to indicate the general approach to measurement.

"Do not be troubled about the matter of administering tests. The instructions for each test are simple and, if followed carefully, should present no difficulty" (169). . . ."*Tests for Basic Interests.* For an investigation into the kinds of activities in which an individual is interested, one of the best known tests is the Preference Record, by G. Frederic Kuder.

"Preferences will certainly be exhibited in the degree of enthusiasm with which an individual enters into various activities. It may be wise to learn the preferences displayed by your successful salesmen as compared with those who fail" (167–68). This last quotation is the complete information given the reader about a test which is recommended for a selection battery!

Sections of the book are devoted to the application blank, and selection of salesmen for small companies. Materials concerned with the application blank are the strongest point in the publication. Even in this section much of importance is ignored. Readers who wish to read further about topics which interest them will be irritated by the author's failure to include sources. The names of writers and researchers are referred to, but the publications and publishers are omitted.

In the opinion of the reviewer, this is a poor book on a topic which is deserving of much better treatment.

Milton E. Hahn

Syracuse University

Traxler, Arthur E. *Techniques of guidance: Tests, records, and counseling in a guidance program.* New York: Harper & Brothers, 1945. Pp. xiv + 394. $3.50.

The first thirty or forty pages of this compact volume may give the reader, sophisticated or otherwise, some blank misgivings. But if he persists, his patience will be richly rewarded. Few books that have come to my attention in recent years have been more appropriately titled or more effective in their progression. From a conventional beginning that toys with definition, staff organization, and costs, the science of guidance gradually becomes the art of bringing into conjunction to their mutual advantage the potentialities of pupils and the resources of school and community.

School and community resources are treated rather generally but with ample reference to sources of information. Understandably the curriculum, which is the school's chief formalized resource, is neither described nor challenged; but the growth of the extracurriculum is given subtle but effective emphasis as a promising adjunct to readin', writin', and 'rith metic.

If facts are the beginning of wisdom, then facts about the pupil—as a learner, a personality, and a social being—are the *sine qua non* of guidance. Kinds of appropriate facts are outlined: home background, school history, mental ability, achievement and growth in fields of study, health, out-of-school experience, interests, special aptitudes, personality, and plans for the future. Use of the questionnaire and the interview is discussed and their values and limitations highlighted.

Tests as instruments of fact finding, however, are given the largest amount of space and the most exhaustive treatment. Aptitude as a concept is intelligently discussed, and likewise the fact of individual differences; but without any attempt at explaining the why. A selected but reasonably exhaustive list of tests carefully and fully annotated is attached, including titles, authors, publishers, research information and references.

An even more extensive array of achievement tests is described with equal fidelity to fact. These fall mainly within the areas of reading, general achievement, English, foreign language, mathematics, science, and the social studies.

Appraisal of personality is treated under two headings: (1) tests, with a rather extensive annotated list included, and (2) anecdotal records and behavior descriptions accompanied with discussions of the values and pitfalls of both tests and records or descriptions.

Planning and administering a testing program and scoring, organizing, and reporting test results are fully and effectively treated with concrete suggestions drawn from actual practice buttressing the general discussion. A variety of services and possible cooperative ventures are pointed out for schools not fully equipped to engineer a program without outside help.

The possible *uses* of test results point up sharply the tremendous lag between test construction and test-giving on the one hand, and on the other, what is done about all this fury of activity. With infinite patience Dr. Traxler describes the instructional uses of tests. He elaborates them so fully and with such reasoned precision that he may feel, as many readers will, that he has argued his case perceptibly beyond the point of diminishing returns. Administrative uses of test results, and by inference, their possible effect upon the curriculum, are it appears "outside the scope" of his book.

Within the conventional setting of school resources, counseling uses of test results are recommended to identify individual weaknesses of pupils, to discover special abilities, and as a basis for conferences on behavior adjustments, course sequences, vocational and educational planning. But tests are only one element in the total guidance program.

"A comprehensive and detailed system of cumulative personnel records is indispensable for the proper functioning of the modern school." Some eighty pages of text and sample forms are devoted to the cumulative record and reporting devices and to dependence on facts of status and progress for immediate or ultimate attainment of desirable educational or career objectives. The ideal is admirable and convincingly presented as within the realm of the attainable. Comparability of test data is projected as essential. Adequate and systematically planned assembling and recording of other kinds of data are stressed as equally important. If both kinds of information are to be significantly useful, they must take their origin not from tests and records as such but from "a study of the nature and purposes of the school and of the pupil." Useful, practical suggestions are given for filing and accessibility of materials.

"The whole thesis of this book, however, is that counselors who are not psychologists are entirely capable of performing the *distributive functions* of guidance and many of the *less involved adjustive functions*, provided they will get acquainted with and use the techniques for knowing individuals." To this end lists of references, not exhaustive, but well chosen, have been appended to each chapter; and the final chapter consists of reading resources for counselors. For those who have the urge to learn their students as well as teach them, Dr. Traxler's volume holds out great promise. For those who are professionals in the field of guidance, the volume is an excellent handbook.

F. S. Beers

Social Security Board,
Washington, D. C.

Scott, Dorothy D., Morgan, Winona L., and Lehman, Ruth T. *Developing a student guidance program in an instructional department.* The Bureau of Educational Research, Ohio State University. 1945. 65 pages.

The monograph describes the procedures used for the past several years in the School of Home Economics at Ohio State University in an attempt to provide more adequate student guidance and to determine techniques useful to the teacher counselor. The plan was inaugurated because it had been found that many students graduated without any faculty member really knowing them, and many made unsatisfactory vocational choices.

The staff accepted the idea that colleges should be concerned for the all-round development of the students, that the guidance point of view should permeate student-faculty relations, and that the college staff can best acquire this through participating in a guidance program.

Responsibility for guidance is divided among three groups. The two coordinators are chiefly concerned with administration although they do some counseling and teaching; the general advisers retain the same advisees throughout their four years in college, hold planned conferences with each of them (and additional conferences upon request) to analyze (the students') progress, plan next steps, and suggest resources for their personal development; and the professional advisers plan policies for selecting students, provide group guidance on professional problems and assist in placement. Most faculty members serve as general advisers and the heads of departments as professional advisers.

Individual Counseling. Each conference with the general adviser is planned with a specific purpose, although plans may be altered if desired. The chief idea is to help students and faculty to get acquainted and to put students in touch with sources of special help which may be needed, such as the Health Service, the Employment Service, and the Psychological Clinic. The first conference is held during Freshman Week, the second in the spring. The sophomore conference comes in the middle of the year and is planned to offer help in making a vocational choice. The last planned conference is arranged for the fall of the junior year and at that time students are helped to analyze what has been accomplished during the first two years and what should be planned for in the last two in order to get the maximum out of college. Students may initiate conferences during the senior year as they may during the preceding years, but no definite conference is scheduled. Informal entertaining of small groups in her office or at the adviser's home provides a very worth while type of contact.

Transfers are assigned to advisers when they enter and they furnish much of the same personnel information which the freshmen do.

Group Guidance. Several types of group guidance are offered. The Freshman Orientation course which is required of all entering freshmen, meets three times a week, and carries two credits. The course includes small group discussions for 30 or 40 students, counseling by instructors and selected older students, collection of personnel information, and administration of most of the tests used. Professional guidance is given through meetings sponsored by professional advisers—sometimes for those specializing in the department and sometimes open to all students. Discussion of pertinent material, student participation in the program, and recognition on the part of the students of the need for the information probably explain the high attendance at these meetings.

The Guidance Office is under the direction of the coordinators and handles much of the routine work of the program, such as providing advisers with all needed forms, keeping confidential folders up to date, and administering tests. The coordinators take charge of scheduling laboratories for underclass students and the professional advisers of those for upperclass students, and in these, students work out their schedules and can obtain individual assistance if it is needed.

Rating of students is done by staff members with a relatively simple form and provision is also made for recording any comments instructors wish to make about the student. Ratings are given on only those points for which the instructor believes she has a basis for judgment and students are allowed to see their ratings when an adviser is available to help them interpret them. Careful self-evaluation is provided for at several stages.

The appendix includes copies of all of the forms used in the guidance program, an outline of the orientation course, and an annotated list of the tests and inventories used.

Clara M. Brown

University of Minnesota

Klein, D. B. *Mental hygiene,* the psychology of personal adjustment. Henry Holt, 1944, xiii + 489.

Mental hygiene has reached the stage where specialists in the field should be factual, critical, and systematic. In the first major section of the text, Dr. Klein fulfills excellently his preface promise to meet these criteria. Here, he discusses in some detail with documentation an objective and eclectic view of the psychoses and psychoneuroses and the prophylaxis pertinent to each of them. The author's second division of the field, called by him "Promotion of Mental Health," is presented in the form of loosely connected essays, mostly on important but selected topics. The home, conscience, motivation, coping with reality, repression, economic barriers, and education are the subjects of the chapters of this section. Despite the fact that there is a voluminous literature on parent-child relationships, only two references appear in the entire chapter on the home. The chapter on education deals with selected important problems but the reader may leave it without realizing the extensive factors in education affecting morale. The influences of the playground, athletics, cliques, and other extracurricular events, for example, are omitted.

The author warns that he will restrict himself "to the confines of mental hygiene territory." This apparently explains the omission of a concise preliminary review of basic psychological concepts as motivation, adjustment, conditioning, inhibition, symbolism, etc., and fundamentals in a theory of personality. It does not explain why psychoses and psychoneuroses are given a disproportionate emphasis at the expense of juvenile delinquency, child and adolescent personality and behavior problems and psychosomatic difficulties.

Dr. Klein's book makes certain distinct contributions to the field. The discussion of specific prophylaxis for each mental abnormality is an example. The forceful chapter devoted to the economic and social aspects of mental hygiene reminds us that there are factors of social control in mental health which are outside of the influence of individual psychologists or psychiatrists. The recent periodical literature is related to a discussion of basic theoretical problems, such as the distribution of mental illness, the extent to which it is constitutional or the result of conflict, the implication of the animal studies and the meliorative and prophylatic roles of the hygienist. These are a reminder that too much has been written in this field without first examining basic premises, as Dr. Klein did.

The task of meliorative mental hygiene as the author sees it is to develop a philosophy of life and high morale. This is not achieved, he stresses, through an array of mental hygiene maxims or drills. This emphasis is a good antidote for some kinds of pseudo mental hygiene. It is congruent with the prevalent non-directive trend in psychotherapy. However, discussion of specific problems and techniques as a means of eliciting insights and building attitudes in youth can be achieved by a skillful presentation without jeopardizing the reader's self direction or independence of action. This reviewer believes Dr. Klein could have strengthened his book pedagogically by greater use of this specific approach. The underclassman, however, should find the text readable, stimulating to insights, and socially progressive.

Fred McKinney

University of Missouri

Bowley, Agatha H. *Guiding the normal child.* New York: Philosophical Library, 1943. Pp. xv + 174. $3.00.

This book which is addressed to parents, teachers, social workers, and physicians contains many practical suggestions for dealing with behavior and adjustment problems of children. Its author, who is a lecturer in the teachers' training college at Dundee, Scotland, has had wide experience with nursery schools and as a psychologist in the Dundee Child Guidance Clinic.

Beginning with an account of the child's development during infancy, the treatment carries forward to the end of the adolescent period. The specific chapter titles are Infancy, The Pre-School Period, Difficulties During the Pre-School Period, The Middle Years of Childhood, Difficulties During the Middle Years of Childhood, Adolescence—Development and Difficulties, and Children and the War.

The volume is exceptionally easy reading and contains numerous short case studies which adds to its interest. It is decidedly not a typical textbook on child psychology, and probably should not be used for this purpose. It would, however, provide excellent supplementary reading in connection with a course in child psychology. It is distinctly a "how" book.

For the most part the suggestions it gives for child guidance appear to be based upon a sound psychology of learning and adjustment. Occasionally, however, the author propounds a theory or takes a position which seems at variance with the best current thought in America. For example, in discussing fears of the pre-school child, she says: "These fears are largely the result of unconscious feelings of badness, and may be termed *anxieties*. . . . He may fear that his parents may desert or harm or starve him if he is unworthy and unlovable because he sometimes harbours bad, hateful feelings towards them. If he feels so to them they may reciprocate and retaliate. . . . This would seem to be at the root of the phobias of the pre-school child which are so very common. Fears of dogs, cats and spiders, for instance, frequently represent a projection of these unconscious fears on to some object in the external world against which he may enlist protection" (p. 63). In other words, the child may be afraid of dogs or cats not because of previous unpleasant experiences with them, but because he has an unconscious fear of his parents. With respect to the development of intellectual traits, the author has a strong hereditarian bias. She says: "The dull child is dull by virtue of his poor mental inheritance. It is important to recognize his degree of dullness, to assess his mental age and so plan a curriculum that suits him and ensures that his needs are satisfied" (p. 92). In view of recent findings with regard to the constancy of the IQ, this seems to be a rather strong statement. Her definition of adolescence as being "that period between 13 and 18 years for boys, and 12 and 16 years for girls" (p. 131) would not generally be accepted as valid in this country. Our boys and girls seldom achieve social maturity or become weaned from their families at such early ages.

Despite these minor criticisms, the book has very much to commend it. It is functional, and highly suggestive with respect to child guidance procedures. The chapter on Children and the War is particularly timely and valuable. Dr. Bowley's book can be read with profit by anyone who desires a straightforward, simple, and dynamic explanation of child behavior problems and methods of dealing with them.

Glenn Myers Blair

University of Illinois

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Studies of teachers' classroom personalities, I. Dominative and socially integrative behavior of kindergarten teachers. Harold H. Anderson and Helen M. Brewer. Stanford University: Stanford University Press, 1945. Pp. 157. $2.00. Applied Psychology Monograph No. 6.

The enrichment of life. Paul N. Elbin. New York 17: Association Press, 1945. Pp. 87. $1.50.

Selection of students for vocational training. Fred M. Fowler. Washington: Government Printing Office, 1945. Pp. 156. U. S. Office of Education, Vocational Division Bulletin No. 232.

Through a dean's open door. A guide for students, parents, and counselors. Herbert E. Hawkes and Ann L. Rose Hawkes. New York: McGraw-Hill Book Co., Inc., 1945. Pp. 247. $2.50.

Principles of guidance. Arthur J. Jones. New York 18: McGraw-Hill Book Co., Inc., 1945 (Third Edition). Pp. 582. $3.50.

Price control and business. George Katona. Bloomington: The Principia Press, Inc., 1945. Pp. 246. $3.00. Cowles Commission Monograph No. 9.

Diagnostic psychological testing. David Rapaport. Chicago 4: The Year Book Publishers, 1945. Volumes I and II $6.50 each, postpaid; set $12.00, postpaid.

The dice of destiny. An introduction to human heredity and racial variation. David C. Rife. Columbus, Ohio: Long's College Book Store, 1945. Pp. 163. $1.75.

Human nature in the making. Max Schoen. New York 3: D. Van Nostrand Co., Inc., 1945. $2.50.

Pioneers of tomorrow, A call to American youth. Hans Weil. New York 17: Association Press, 1945. Pp. 81. $1.25.

Frontier thinking in guidance: An anthology of significant thought in the field of guidance. Edited by John R. Yale. Chicago 4: Science Research Associates, 1945. Pp. 160. $2.00.

Practical handbook for counselors. Developed by New York State Counselors Association. Chicago: Science Research Associates, 1945. Pp. 160. $1.50.

testing. The second problem is a matter for mere arbitrary agreement among qualified psychologists. If it is agreed that in the total intelligence score the attention test is to have a value of one third or some other value, well and good. But in the end we will have an agreed-upon series of tests measuring an agreed-upon ability or group of abilities.

One may well pause to wonder whether the term "intelligence" as it is now (apparently) used is of any value to psychology. At present one's intelligence is merely a relative ranking on a battery of tests of the higher central processes. As such it represents a summation of the scores made upon the various component tests in the system. At best it is a sort of abstraction or "average" intellectual level. A psychograph, representing an individual's levels on the various tests, would be much more elucidating than an intelligence score which merely hides relative strength and weakness in the component abilities.

VARIOUS "KINDS" OF INTELLIGENCES

Again taking their cue from sociology and practical life some psychologists have suggested and introduced various "kinds" of intelligence. Thorndike has advocated these various species of intelligence as a logical outcome of his view that intelligence is but a bundle of specific and independent capacities. But it appears merely to add more confusion to an already sufficient confusion. If we speak of abstract, concrete, manual, and social intelligence then by virtue of the fact that we use intelligence throughout it is suggested that there is an intelligence common to all forms of activity. The thing then works very nicely over into Spearman's common factor view. Would it not clarify matters if we spoke of just one kind of intelligence, possessed in different degrees by different people, and proceed to define intelligence in agreed-upon psychological processes? It would not be necessary to know the ultimate nature of intelligence.

EFFICIENCY RIGHT NOW VERSUS "CAPACITY"

Any test measures what the subject can do right now, just as he is at this moment and under the external circumstances which are present. We may call this "present efficiency." But is it necessarily the man's "capacity?" Capacity should mean just what the term implies, namely, the person's ability under the optimum conditions for that organism. Capacity is apparently being used in two different senses, one of which should be entirely dropped or else very clearly defined. A person's true capacity could be measured only when that capacity has reached its maximum maturation and practice. The intensity of effort and interest must be maximal, to say nothing of the many other favorable external and internal conditions which would be pre-supposed. Capacity is, at best, an inferred thing and its inference is fraught with much chance for unfairness.

Perhaps it would be well to speak of measuring one's "efficiency" when we are seeking a person's demonstrable capacity as it reveals itself at the time of the testing. The term "efficiency" would then mean merely what the subject was able to do at that time. It would have no implications beyond this. The term "capacity" could then be used to signify what the person *might* do if his training, attitude, and physical condition, were *most favorable* for the most complete development of the given capacity or ability. A child reveals an intelligence quotient of, say, 100 on the Stanford Binet tests. Later on by systematic physical treatment, good food, or training the intelligence quotient has been raised to 135.[1] We may well say that the test did not get at capacity but merely efficiency at the time. To what extent we may infer capacity from efficiency, in the case of any test whatsoever, can be determined only by careful study.

[1] Just such a case has been recently reported by June E. Downey, Journal of Applied Psychology, Vol. XI, 1927, page 52. Other such cases have come to the writer's attention.

A STUDY OF THE RELIABILITY OF MENTAL TESTS AS USED IN OKLAHOMA CITY

C. L. Cushman

Oklahoma City Public Schools

This study is given as a contribution to the available information concerning the reliability of mental tests when given by class room teachers.

In the Oklahoma City schools it has been the policy for the past four years to give each child entering the kindergarten or first grade a Binet Intelligence Test. These tests are given at about the time of entering the first grade. The plan in Oklahoma City has been to train all of the primary teachers in the giving of individual tests and to have each teacher responsible for testing the pupils of her own room. In the intermediate grades pupils are retested with the Otis Self-Administering Test of Mental Ability.

For a number of years the permanent record cards in the schools were not forwarded from school to school with the transfer of children but each school made out a new card for every entering child regardless of where he last attended. The data for this study were assembled from these duplicate permanent record cards which are now in the files of the Research Office.

The time intervening between the two tests ranged from one month to more than two years with an average interval of between eleven and twelve months.

It is true, of course, that in certain respects the children represented in this study are not a typical group. That is, they are drawn from the transient element of the city's population. The fact that the average IQ., derived from the 352 Binet tests, included in these figures was 100.3 and the average of the 168 IQ.'s derived from the Otis tests was 95.4 would seem to indicate that the group was not an inferior one. It is entirely possible, however, that other factors enter in to make mental tests less reliable with this type of children.

Table 1 gives the main facts relative to the reliability of the tests as shown by this study. An examination of this table will show that as used in Oklahoma City the Binet and Otis tests are equally reliable.

Table 2 is built to show the tendencies of the IQ. within the different portions of the total distribution. In this table only the Binet tests are considered. This table shows clearly two things; first, the reliability of the test results increases rapidly as the scores approach the mean of the distribution; second, where an initial test gives an IQ. greater than

TABLE 1

	TESTS COMPARED			
	Binet and Binet	Binet and Otis	Otis and Otis	First and Second
Number of cases	144	54	62	260
Average difference in I.Q. between first and second tests (disregarding sign)	7.8	9.5	7.2	8.0
Coefficient of correlation between the two tests	0.74	0.72	0.78	0.74
P.E.$_r$	0.025	0.041	0.032	0.019
P.E. score*	5.1	5.8	4.5	5.1
Distribution according to number points difference in I.Q. Number of points difference:				
0–10	70.1%	62.5%	71.1%	67.7%
11–20	23.6%	31.3%	26.9%	26.9%
More than 20	6.3%	6.2%	2.0%	5.4%

* P.E. score = $0.6745 \sqrt{1-r^2} \times \sigma$.

TABLE 2

Tendencies of Binet Scores according to position of the initial test in the distribution

	SCORES ON INITIAL TEST								
	50–59	60–69	70–79	80–89	90–99	100–109	110–119	120–129	130–139
Number of cases	1	4	8	22	32	37	20	13	7
Average of first tests	59	64	73.6	81.7	91.8	104.4	112.0	122.4	131.7
Average of second tests	66	73.5	83.4	87.5	95.7	100.7	108.0	114.5	110.7
Difference in the two averages*	−7	−9.5	−9.8	−2.8	−3.9	+3.7	+4.0	+7.9	+21.0
Number of cases in which first test was higher than second	0	1	2	8	8	17	13	11	6
Number of cases in which first test was lower than second	1	3	5	11	16	12	6	0	1
Number of cases where scores were the same	0	0	1	3	8	8	1	2	0
Average disagreement between two tests	7	10	10.3	7.9	5.6	6.4	7.4	7.9	22.4

* + indicates average score of first tests was higher than average of second tests by amount indicated. − indicates average score of first tests was lower than average of second tests by amount indicated.

the mean the chances that this IQ. is too high are much greater than the chances that it is too low, while in those cases where the initial test gives the IQ. less than the mean the reverse is true. In both cases the distance by which such scores are probably too high or too low increases with every increase in distance from the mean. It will be noted that in spite of the small number of cases in some of the outside columns there is no exception in this tendency to regress toward the mean anywhere in the table.

An examination of the two tables would seem to indicate that with a testing program organized such as Oklahoma City's little importance should be placed on the mental scores of the extremes of the distribution.

THE WILL-TEMPERAMENT OF INDIANS[1]

THOMAS R. GARTH
University of Denver
AND
MARY A. BARNARD
Colorado Psychopathic Hospital

In these modern days the term racial temperament has a very well established usage. We seem to see in the Chinaman one sort of temperament, in the Negro another and in the Indian another to say nothing of Italian temperament and Irish nature. It would be hard to shake the faith of latterday members of society in civilization away from such a belief. Somehow we have come to think just in this way and we have not asked ourselves the question if these temperaments are after all really racial, that is, native and not acquired. It is time that psychology were seeking to answer the question as to whether or not there are such things as racial temperaments.

Efforts have been made to study the temperament of the southern Negro scientifically (1), (2) and the device used for the purpose was either altogether or in part the Will-Temperament Test devised by Dr. Downey. The results of these studies, one by McFadden and Dashiell in North Carolina and the other by Sunne in New Orleans are indicative but not conclusive.

The writers recognize the inadequacy of all tests of personality used so far in making studies. This would include of course the group test devised by Dr. Downey which we are using to study Indian personality but we believe, with others, that it is the most promising attack of temperament yet made and that it bids fair to have a wide application as it is more and more perfected.

---

[1] Paper read before Section II, American Association for Advancement of Science, December 29, 1926, at Philadelphia, Pa.

PROBLEM

The problem here is:

1. To obtain, if possible, reliable measures of full-blood Indians for the various tests of the Group Will-Temperament Test.

TABLE I

Scores Obtained in Group Will-Temperament Test by 101 Whites and 170 Full-Blood Indians

	TESTS												TOTAL SCORE
	A	B	C	D	E	F	G	H	I	J	K	L	
For white subjects													
Med.....	4.11	5.12	4.15	6.10	5.23	5.03	5.03	7.08	3.71	5.55	5.42	3.88	59.54
Q........	2.05	3.25	1.82	2.17	1.12	2.65	2.47	2.69	1.53	1.91	2.66	2.13	5.39
$\frac{SD}{Med}$.....	0.37	0.60	0.34	0.30	0.20	0.48	0.44	0.48	0.28	0.17	0.48	0.39	0.99
For full-blood Indian subjects													
Med.....	2.83	3.08	4.59	3.10	5.31	2.98	3.44	6.10	6.85	5.90	1.95	9.68	55.86
Q........	1.20	3.59	1.33	1.47	1.53	2.34	2.97	2.76	3.01	2.59	0.95	0.88	2.25
$\frac{SD}{Med}$.....	0.17	0.51	0.19	0.20	0.22	0.33	0.40	0.39	0.43	0.37	0.12	0.12	0.31
Overlaps*...	22%	47%	50%	18%	58%	35%	46%	48%	72%	57%	16%	24%	34%

* Per cent of full-blood Indians attaining and exceeding the median of the Whites.

(A) Speed of Movement; (B) Freedom from Load; (C) Flexibility; (D) Speed of Decision; (E) Motor Impulsion; (F) Self-Confidence; (G) Non-Compliance; (H) Finality of Judgment; (I) Motor Inhibition; (J) Interest in Detail; (K) Coördination of Impulses; (L) Volitional Perseveration.

2. To compare these with similar measures of white individuals so as to ascertain if the groups behave differently when so tested.

SUBJECTS

There were two groups of subjects. One was a group of 101 White students in a High School of the City Schools of Denver. All but 5 of these were from the eleventh and twelfth grades and all but 7 were from seventeen years and on in age. Of these 7, one was fifteen years and the others sixteen years old. The sexes were about equally divided there being 49 girls and 52 boys. The Indian group was composed of 170 full-blood Indians from the United States Indian Schools at Santa Fe and Albuquerque, New Mexico, and Rapid City, South Dakota. There were 85 boys and 85 girls. About half of the group were full-bloods from sedentary tribes being nearly altogether Pueblo Indians; the other half were from different nomadic tribes, mostly Navajos, Apaches and Sioux.

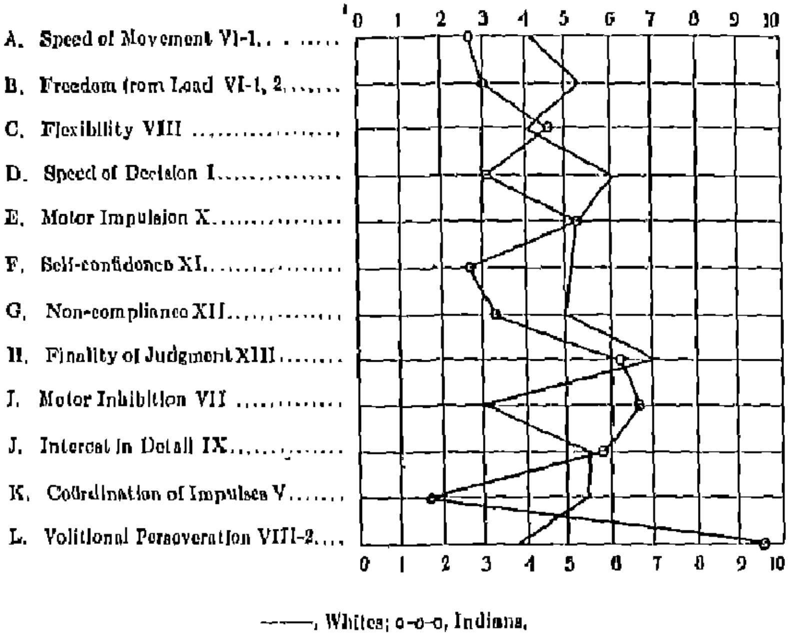


FIG. 1. SHOWING WHITE AND INDIAN GROUP PROFILES

Administration and scoring

The accepted technique of administration of the group test was followed as closely as possible although it is realized that the technique is very difficult of execution. There was much practice by the experimenters in administering the test before the tests were given from which the data in table 1 were obtained. In scoring the tests the writers had some valuable assistance from the author of the tests for which they make acknowledgment. Regardless of these safeguards it is felt that variations found in these results may be partly due to variations in technique and scoring.

HANDLING THE DATA

In table 1 will be found the Medians and Quartiles of the 101 White subjects and of the 170 full-blood Indian subjects for each of the separate test items and for the total score. There will be found also measures of overlapping of the Indian measures on the white medians. Figure 1 gives the group profiles of the two racial groups.

INTERPRETATION

In the first place it will be seen from examining the medians of both racial groups and the profiles that there are points of similarity as well as difference; this is true on a basis of mere medians. But it will be noted that in some of the tests in both racial groups there is high variability and that therefore such medians are somewhat unreliable. This is true in eight of the tests for the white group and six of the tests for the Indians.

If one takes into consideration the overlapping of the Indians on white medians it will be found that the racial differences indicated by the medians in tests for Speed of Movement, Speed of Decision, Motor Inhibition, Coördination of Impulses and Volitional Perseveration indicate real differences as found between the two racial groups.

The overlapping for total score also indicates a tendency toward a real difference in that respect. That is to say, the test indicates that the Whites are somewhat more speedy than the full-blood Indians in so far as the test indicates what it claims to test. They are just the same in flexibility but much more inclined to make quick decision than are the Indians. In motor impulsion the two racial groups are just about the same but the Whites show a greater amount of self-confidence than do the Indians. In non-compliance the results for the two groups are not very conclusive though the Whites are slightly stronger than the Indians it would appear upon comparing mere medians. The same may be said of finality of judgment although there is more similarity in the two group measures and the results appear more trustworthy, i.e., both groups appear rather strong in this respect. In motor inhibition the Indians certainly have the better of the Whites though they are slightly more variable. In fact the test indicates that this is a rather strong trait in the Indian group and the weakest trait for the Whites. In interest in detail the racial groups are again rather similar though the Whites are less variable. There is a striking weakness in coördination of impulses in the Indians as here measured. In this respect the Whites are (according to the author of the tests) just about normal. The difference seems to be genuine as measured. On the other hand the Indians attain a very high score in volitional perseveration and the variability is excessive under the circumstances whereas the score of the Whites is much lower and their variability is high. If we may take the total score as indicating strength of personality the Whites would appear to have stronger personalities than the Indians.

COMPARISON WITH NEGROES

Unfortunately the tests given to these Indians were not the same as those given by the investigators mentioned above but we may make a rough comparison. MacFadden and Dashiell conclude that the Negroes tested by them were slower than their White group and in this respect the findings for the Indians are similar in that they are slower than the Whites. They also found the Negroes to be quicker in the matter of speed of decision than the Whites and we find the Indians slower than Whites. These investigators found the Negroes more self-confident but we find the Indians less so than the Whites. The Negroes were stronger in motor impulsion than Whites but the Indians are here found to be similar as measured. The Indians and Negroes seem to be slightly similar in being weak in non-compliance as compared with Whites. The Negroes are weaker in motor inhibition than the White group with which they were compared but here the Indians excel the Whites. Both the Negroes and these Indians proved weaker in coördination of impulses than the white groups with which they were compared and in volitional perseveration the Indians excel Whites while the Negroes fall below them in this respect.

The last finding of excellence in volitional perservation is supported by a finding by Garth (4) in his study of White and Indian and Negro work curves where the Indians held out better in a continuous task in addition than either Whites or Negroes although the Indians were nomadic Indians. His findings for sedentary Indians showed them as not excelling in the work curve test (5).

SUMMARY

When Indians are tested with the Group Will-Temperament Test and the results are compared with results from testing Whites of similar age we find:

1. The comparison of the total score indicates that Whites possess stronger personalities than Indians as here measured.
2. While the measures of various traits in both racial groups show great variability those showing least variability are: For Indians, Speed of Movement, Flexibility, Speed of Decision, Motor Impulsion, Coördination of Impulses, and Volitional Perseveration; for Whites, Flexibility, Motor Impulsion, Motor Inhibition, and Interest in Detail.
3. The greatest differences between the two racial groups as indicated by measures of overlapping are to be found in Speed of Movement, Speed of Decision, Self-confidence, Coördination of Impulses, Volitional Perseveration and Motor Inhibition.

4. Interpreted the above means that the Whites are more speedy in the test than are the Indians, they are quicker in making decisions, they are rather more self-confident, they have less motor inhibition, better coördination of impulses and are poorer in volitional perseveration.

5. The greatest differences as found here between the races are in speed of movement, speed of decision, coördination of impulses and perseveration. In only the last of these does the Indian excel.

6. It is rather dangerous to venture any conclusion with reference to comparison of Negro and Indian Temperament but it would appear that Indians are less speedy in decision than are the Negroes and they appear to be more given to motor inhibition than the Negroes.

REFERENCES

(1) McFadden, J. F., and Dashiell, J. F.: Racial differences as measured by the Will-Temperament Test. Jour. Applied Psy., 1923, vol. 7, 30-53.

(2) Sunne, Dagny: Personality tests: White and Negro adolescents. Jour. Applied Psy., 1925, vol. 9, 205-280.

(3) Freeman, F. N.: Mental Tests, 1926, 101–303.

(4) Garth, T. R.: Racial differences in mental fatigue. Jour. Applied Psy., 1919, vol. 5, 235–244.

(5) Garth, T. R.: Mental fatigue of Indians of nomadic and sedentary tribes. Jour. Applied Psy., 1926, vol. 10, 437–451.

THE FOURTH INTERNATIONAL CONGRESS OF PSYCHOTECHNIQUE

PORTIA HOLMAN

Official Interpreter and Assistant to Dr. Lahy

Between the 10th and 14th of October, 1927, the fourth International Conference of Psychotechnique was held at the Institute of Intellectual Coöperation of the League of Nations in Paris. This society was founded in 1920 by Professor Claparède (Geneva) with the aim of uniting at conferences and by other means all those whose work and interests are in psychotechnique.

Its conditions of membership are very elastic. There are no fees and no articles to subscribe to; its chief purpose being to unite as many psychotechnicians as possible so as to be truly international and representative.

The Paris Conference was the first since 1922 and by the great increase in attendance as in other ways reflected the development that has taken place in this science in the last five years.

The wishes of its founders that a large body of scientists representative of a large number of nationalities should be brought together were fully realised. In particular it was the first time delegates from the United States of America or from the United Soviet Republic of Russia had been present. Consequently it was the first meeting which can be considered as truly representative of the trend of psychotechnical development in all parts of the world.

If that trend may really be judged by the views of the representatives at this conference, the meeting had a special significance in marking the stage at which the science of psychotechnique has now arrived.

As Mr. Lahy expressed it in his report the conference appeared as a "public and international examination of conscience."

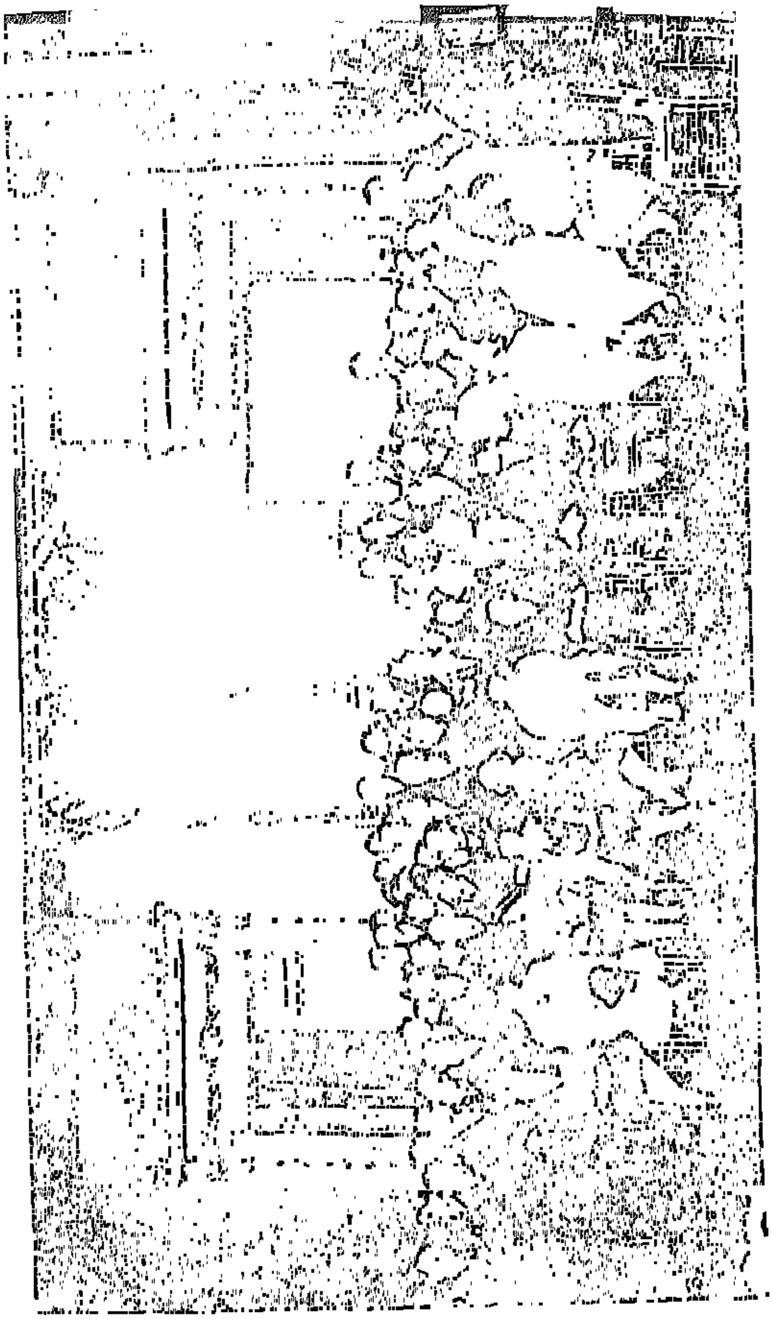
This important meeting seems to have been in accord with a general desire on the part of psychotechnicians in all countries to get into closer touch with one another, to find out more clearly what is going on elsewhere and to create harmonious coöperation between the researches and investigations of different countries.

There was a very general feeling that the relative isolation of psychotechnicians in any one country and their ignorance of what was taking place elsewhere had led to too great a multiplication of effort in the same field and a loss of valuable comparisons because in different places investigators had worked on the same material by similar but not strictly identical methods. Nevertheless one of the most striking conclusions to be drawn from the conference was that, notwithstanding the almost complete lack of coöperation in the past, workers in different countries had arrived independently at the same methods and ha the same ideas as to the broad and general lines of the science.

At the conference there was a definite unity given by the universal interest in general questions of method, control an interpretation. No matter what topic was being discusse almost invariably it was to this aspect of it that both the paper read and the subsequent discussion were directed. Indeed, th whole question of method and of the place of mathematics an statistics in psychotechnique had been given special prominenc in the program. From this discussion two outstanding problem arose: (1) the use and reliability of the correlation coefficient (2) the question of the constancy of subjects, i.e., the reliabilit of a single test as the measure of the subject's capacity.

This second question which clearly affects the whole procedur of psychotechnical investigations aroused the demand for "gigantic international experiment" as one speaker well put i

Among the discussions of particular subjects one might stres the question of tests of character as being of special interest to th conference. There was a general agreement that for purpos of vocational guidance and selection existing tests are far fro adequate. It was felt to be most urgent that attention should directed to the problem of supplementing them both by those moral qualities and by tests of the higher mental qualities, f example, organizing ability.



HOSTESSES AND SOME OF THE GUESTS AT THE AFTERNOON TEA

The headings under which discussion took place were: Principles and Method of Psychotechnique, The Educability and Constancy of Subjects; The General Question of Tests, The Problems of Training (of special skill), The Relation of Psychotechnique to Education, The Relation of Psychotechnique to Psychiatry, The Question of Selection with Special Reference to Transport Services.[1]

The one unsatisfactory feature of the conference was its brevity. It is, of course, of the nature of a conference that it should last only four or five days but it was particularly regrettable in this instance because the number of papers was exceptionally large. The problems arising out of them might profitably have been discussed for a year.

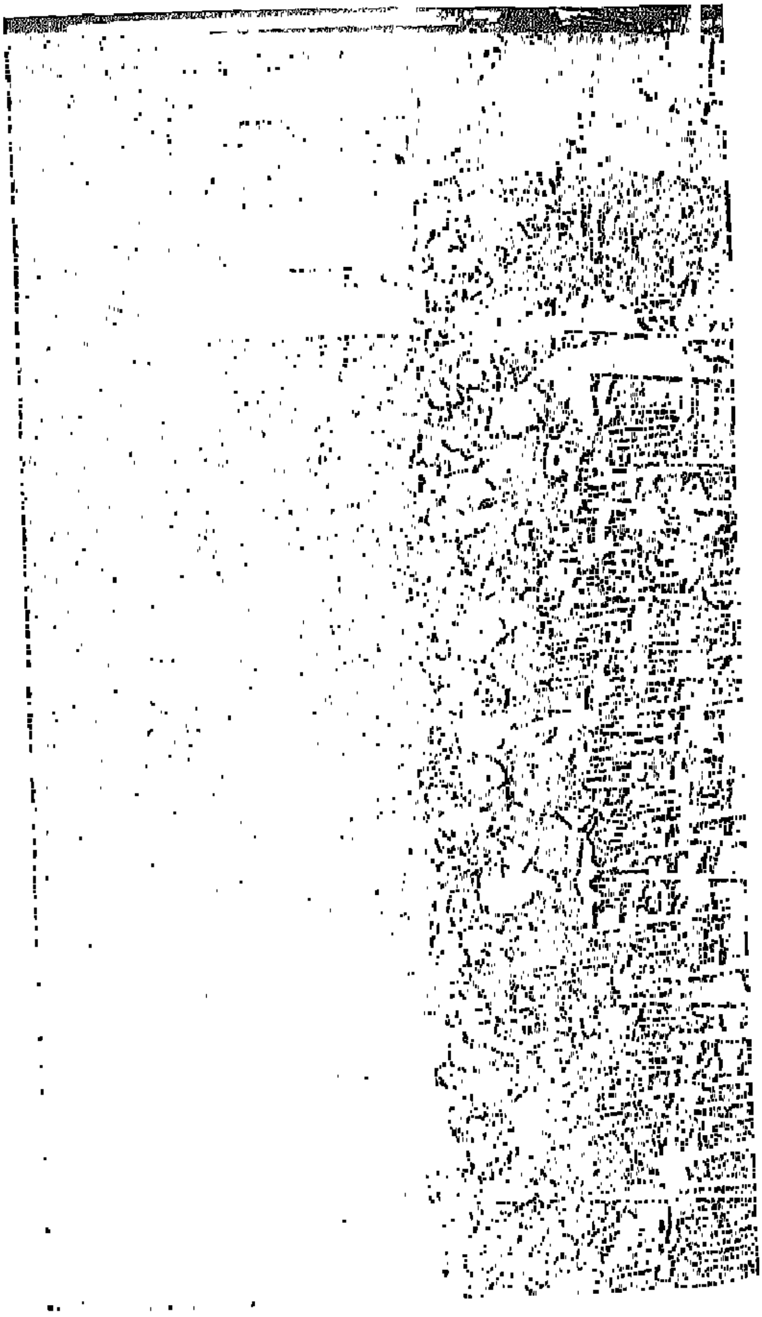
The number of papers was so great that many will be communicated for the first time in the proceedings of the conference to be published shortly. Of those which were read time permitted the discussion only of a small number and frequently with a time limit.

Differences in languages and the lack of time for complete translations presented another obstacle to exchanges of views. At least there was no lack of private discussion between the crowded sessions and, as we have noted above, in spite of material difficulties the general trends of ideas were very clearly manifested.

Most of the members present were in agreement with the views put forward by Mr. Wilson (London) in a special paper dealing with the organization of the association. He argued that the association should be more than a mere agency for holding occasional conferences and that as far as possible continuous investigations of an international character should go on in the inevitably lengthy intervals between conferences. In particular the association might usefully perform the function of codifying and bringing together an

---

[1] The papers on this subject were read at a special meeting of the psychotechnical laboratory of the Société de Transports en Commun de la Région Parisienne in which members were able to supplement discussion by an inspection of the practice in this laboratory.



The First Session of the Congress

authoritative statement of the work that was being done in each field in every country so that even without deliberate collaboration psychotechnicians in different countries might be in a position to pursue the same ends by the same means.

In view of the general agreement that psychotechnical experiment can in general give valid results only when they are made on very large numbers of subjects it was urged that all possible steps should be taken to facilitate the making of strictly comparable investigations in different laboratories.

It was also felt that certain subjects might best be investigated by an international commission. In response to this feeling topics for discussion by a number of commissions were put forward and for those on which the conference agreed a president of the commission was elected. The conference delegated to each president the task of choosing the commission's personnel and that of determining the manner in which it will work.

The Commissions are as follows:

1. Psychotechnical Study of Accidents. Professor Marbe (Wurtzburg).
2. The influence of the Environment on the Worker. Mr. Wilson (Industrial Fatigue Research Board, London).
3. Study of Educability. Professor Claparède (Geneva).
4. The Influence of the Exertion of the Worker on Output. Professor Lipmann (Berlin).
5. The Practical Measures to be Taken to Form a Complete Collection of Existing Tests, with Descriptions of Methods and Results so far Obtained. Professor Piéron, (Paris).
6. An Attempt to Unify the Psychotechnical Vocabulary. Madame Baumgarten (Soleure, Switzerland).

One or two important steps were taken with regard to the organization of the association itself. At a preliminary meeting before the official opening of the conference a tentative plan of fusion of the Association of International Conference of Psychotechnique with the International Association of Psychology and Psychotechnique was put forward. This fusion was voted on in the opening session and it was subsequently decided that the body arising from the fusion should henceforward be known as the International Association of Psychotechnique.

A general "*committee of direction*" of this association has been set up composed of not more than two representatives of each country.[2]

The Institute of Intellectual Coöperation has placed rooms for the Secretariat-General permanently at the disposal of the association.

Dr. Moeller of Riga has been asked to undertake the direction of an Information Service, by which it is hoped to keep the *General Secretariat, and through it members of the association,* in constant touch with developments in every country.

The town of Utrecht has invited the association to hold its next meeting there in September or October, 1928. This invitation was accepted by the members of the Fourth Conference but the exact date was not decided.

The proceedings of the conference are being published and members will be informed as soon as possible of the price at which the volume is available.

COMITÉ DIRECTEUR DE L'ASSOCIATION DES CONFÉRENCES INTERNATIONALES DE PSYCHOTECHNIQUE

Membres

Baumgarten-Tramer, F. (Mme.). Docteur en philosophie. Soleure (Suisse).

Bingham, V. *Directeur de Personnel Research Fédération, président* de l'Association américaine de psychologie appliquée. New York.

Christiaens, A. G. Directeur de l'Office intercommunal d'orientation professionnelle de l'agglomération bruxelloise.

Claparède, Ed. Dr. Professeur à l'Université de Genève.

Corberi, Dr. Asilo de Montbello près Milan.

Decroly, O. Professeur a l'Université de Bruxelles.

Ferrari. Professeur à l'Université de Bologne.

Grunbaum, A. A. Dr. *Docent* de Psychologie à l'Université d'Amsterdam, rédacteur de "l'Administrative Arbeid," chef du laboratoire psychologique à la clinique Valerinsplen.

Joteyko, J. (Mlle.). Professeur de physiologie à l'Université de Varsovie, présidente de la société polonaise de psychotechnique.

Lahy, J. M. Directeur du laboratoire de psychologie appliquée à

---

[2] The list of members of this committee is appended.

l'Ecole pratique des Hautes Etudes et à l'Institut de psychologie de l'Université de Paris.

Lipmann, O. Directeur de l'Institut de Psychologie appliquée à Berlin.

Miles, G. H. Directeur-adjoint à l'Institut national de psychologie industrielle de Londres.

Mira, E. Directeur de l'Office d'orientation professionnelle à Barcelone.

Moede. Professeur à l'Ecole technique supérieure de Charlottenburg, Directeur de l'Institute de psychotechnique industrielle de Charlottenburg.

Myers, Ch. Directeur de l'Institut national de Psychologie industrielle d'Londres.

Piéron, H. Professeur au Collège de France et à l'Institut de Psychologie de l'Université de Paris, Directeur du Laboratoire de Psychologie physiologique à la Sorbonne.

Roels, E. Dr. Professeur à l'Université d'Utrecht.

Rossolimo, G. Professeur de neuropathologie; Directeur de la Clinique pour maladies nerveuses et de l'Institut neurologique à la 1ère Université de Moscou.

Spielrein, I. Professeur, chef du laboratroire psychotechnique de l'Institut pour la protection du travail à Moscou.

Wojciechowski, J. Ingénieur Directeur de l'Institut Psychotechnique de l'Ecole du Bâtiment et du Bureau des épreuves psychotechniques des chemins de fer à Varsovie.

Secretaire General, J. M. Lahy. Paris.

President (jusqu'a la prochaine conférence). Dr. Toulouse, Ed. médecin en chef directeur de l'hôpital Henri Rousselle, directeur de l'Institut de psychiatrie et de prophylaxie mentale à l'Ecole des Hautes Etudes, Paris.

NOTES AND NEWS

The Fourth International Conference of the New Education Fellowship met at Locarno August 3 to 15, 1927. The English Report of this World Conference is now ready. Under the general theme, "The True Meaning of Freedom in Education" are the sections: Psychological Freedom, Freedom through Method, through Environment, through Co-Education, through Creative Art and through International Understanding. Among the speakers present were Dr. A. Adler, Prof. P. Bovet, Dr. O. Decroly, Mrs. B. Ensor, Dr. E. Rotten, Dr. A. Ferribière, Dr. Del Manzo, Dr. H. Rugg, Dr. Carleton Washburne, Dr. Lucy Wilson and Sir Jagadis Bose. The price of this English Report is 60 cents. Agents in the U. S. A.—Bookshop for Boys and Girls, 270 Boylston Street, Boston, Mass.

On November 17–18, 1927, under the auspices of Personnel Research Federation a conference dealing with the human factors in industry was held at the Hotel Pennsylvania, New York City. The aim of this conference was to promote scientific investigations of the following: Susceptibility to Accidents and Accident Reduction, Measurement of Leadership, the Personal Interview: Its Technique and Reliability in Fact-Finding, and the Discovery and Measurement of Mechanical Abilities. At the morning sessions there were papers and reports on researches by members of the Federation. In the afternoons groups were formed for round-table discussion on the topics given above. The chairman of the four groups in the order given were Mr. Eugene J. Benge of the Philadelphia Rapid Transit Company; Dr. F. H. Allport, Syracuse University; Dr. Bruce V. Moore, Pennsylvania State College, and Dr. M. S. Viteles, University of Pennsylvania. Col. Walter V. Bingham, Director, Personnel Research Federation, who with Mrs. Bingham traveled in Europe this fall and attended the meetings of the Fourth International Congress of Psychotechnique in October in Paris, will give a paper on Recent Developments in Industrial Psychology in Europe.

Child Labor Day will be observed for the twenty-first year the last week in January, 1928. To make known the facts concerning child labor and to stir the public to demand their correction is the purpose of Child Labor Day—January 28 for synagogues, January 29 for churches, and

January 30 for schools and clubs. The National Child Labor Committee, 215 Fourth Avenue, N. Y., will send free of charge to any interested persons or groups an analysis of the law of their state and other material to aid in the observance of Child Labor Day.

Leland Stanford University announces that Dr. Robert H. Seashore, National Research Fellow in Psychology, has constructed a portable outfit for measuring sensori-motor coördinations. According to Prof. W. R. Miles this apparatus is to be called the Stanford Motor Skills Unit. It includes six tests which make use of the same mechanisms in several different ways. The integrated scores are all read from counters. A motion picture film (10 minutes) has been prepared showing the unit being demonstrated and used.

"The Classification of Pupils in Private Schools." The recent report of the results of a research program conducted by twenty-six schools of The Private School Teachers' Association of Philadelphia throws into sharp contrast the scholastic standards of public schools and those of private schools. The report explains why the same sets of standards cannot be used for both, and sets forth norms for aid in the proper classification of private school students. The sixteen tables included in the report provide norms and ranges for the various grades in chronological age, mental age and educational age. The procedure in using them is described and examples furnished. Price 25 cents a copy. Requests should be addressed to R. C. Clothier, Chairman, Haverford, Pa.

The International Library of Philosophy, Psychology and Scientific Methods has made arrangements with Dr. Horace B. English to bring out a Student's Dictionary of Psychological Terms. It is planned to treat from 2000 to 4000 terms, including particularly those which students are apt to meet as soon as they read beyond elementary textbooks. The work is designed as a defining dictionary, systematically recording actual usage rather than attempting to legislate. The compiler will be very grateful for any suggestions received by the middle of February and addressed to him at Antioch College, Yellow Springs, Ohio.

BOOK REVIEWS

A. Myerson, M.D. *The Phychology of Mental Disorders.* Macmillan Company, 1927. 135 pp.

This little book is somewhat mis-titled, as the point of view of the author is throughout avowedly that of the psychiatrist and not of the psychologist.

The book is written in a style easily comprehensible to the layman. Indeed, it is for the layman that the book is intended. Myerson describes various ones of the ordinary mental disorders, telling something of the underlying causes so far as they are known, and suggests what should be the social attitude in regard to the psychiatric patient. His whole discussion seems to be eminently sane and practical.

While admitting that psychiatry may be of great aid in solving the problem of crime, he insists that psychiatry cannot do it all and that the problem is mainly a social one and not one for the physician.

The book is well worth the reading.

Edward Andrews Lincoln. *Sex Differences in the Growth of American School Children.* Warwick and York, Inc., 1927. 109 pp.

In this little volume, Lincoln has gathered together and summarized practically all of the investigations which have been made on the problem of sex differences of school children. It is to be noted that he limits his discussion specifically to children of school age, and his findings can be extended to adults only as tentative projections.

In general, Lincoln's results are negative; that is, differences between the sexes are not found or are found in such slight degrees that after taking into account the marked over-lapping, no adequate adjustment as to sex is possible. The above statement does not apply to physical growth, in which respect girls are found to be in general physically more mature in every way than boys of a corresponding age. But in intelligence, in school accomplishment, and in variability no significant differences are discovered. The problem of school adjustment is then a problem of adapting to individual differences without respect to sex.

D. Starch. *Educational Psychology.* Revised Edition. Macmillan Company, 1927. 568 pp.

To the one who is already acquainted with Starch's earlier book, little more comment concerning the present volume is needed than to say that it is written from very much the same point of view and has been brought on the whole very well up to date. The one particular addition is a very brief chapter on Mental Hygiene, a chapter which is probably too brief to be of very great value. In the light of the recent discussions concerning instinct, many would have liked to see the subject treated a little more broadly. Starch's own belief in the uselessness of the discussion, however, is of course warrant toward his neglect of it.

Most of the recent investigations of note in Educational Psychology have found their way into the book. The reviewer feels that the volume would have been of more practical value if some of the standard tests used in the investigations relating to school subjects were of more recent date. In many instances references are made to tests which have long since been supplanted. It is of course a stupendous task to keep up with all of the new tests which are being devised, and a book of this sort should not be expected even to list all of them. However, the statement that "for measuring general merit in written composition *two* scales have been constructed" seems hardly up to date (Italics mine).

As in the earlier edition, a large proportion of the text, approximately forty per cent, is given to the psychology of school subjects, thus reducing the value of the book for a department in which a separate course is given on these topics. For a general course, however, this probably represents the best book available.

D. A. WORCESTER,
University of Nebraska.

JANE HILLYER. *Reluctantly Told.* Introduction by Joseph Collins, M.D. Macmillan Company, New York, 1926. 205 pages.

Miss Hillyer, the sane, intelligent person, is able to stand aside from herself, as it were, and view Miss Hillyer, the psychopath, almost as objectively as though she were viewing someone else. But the sane Miss Hillyer remembers so vividly many of the mental experiences of the insane Miss Hillyer that she is able to give the reader the subjective as well as the objective view. This is a rather awkward way of saying that she is able to describe the experience of going crazy without indulging in self-pity or hysterics about herself. Her account includes much that the serious student of mental pathology would wish to know. Something of her heredity, much of her childhood and early home environment, the breakdown, her hospital experience, long periods of oblivion, an attempt at suicide, returning interests and activities, the return to her own world while yet but precariously balanced, and the long hopeful struggle that, one infers, is still going on to remain normally healthy minded. The writer displays a keen insight into the causes of her trouble, and more particularly the factors entering into her restoration to mental health.

She has a very unusual and very interesting story to tell. She tells it beautifully, simply, and with a touching restraint. Neither confessing nor boasting, she merely relates, and it is in this objective viewpoint that much of the value of the work lies.

In his introduction Dr. Collins, besides assuring the reader of his belief in the genuineness of Miss Hillyer's experience and discussing some of the salient points of her case, presents the psychiatrist's view of how mental abnormality should be regarded.

AMOS C. ANDERSON,
Ohio University.

E. JORDAN. *The Life of Mind.* Charles W. Laut & Company, Indianapolis, Ind., 1925. 324 pp.

Jordan's "Life of the Mind" develops into little more than the author's own rather narrow philosophy of a non-scientific life. The general trend of the book is that belonging to an era that had been buried, supposedly, for at least a decade.

Although following topics that are closely allied to those found in any elementary textbook of psychology, the author disdains to mention practically all of the usual hypotheses except in a most derogatory way, —not for any constructive reason, but seemingly merely because he is violently opposed to anything that may be termed "scientific." In spite of this stated viewpoint we find him turning to the scientific discoveries of physiology and other sciences when an explanation of sensations is desired.

The section devoted to auditory sensations is typical of the whole book. Noises and tones are to be distinguished not because of different characteristics of vibration rate, but merely from the assumption that the human being "likes" tones and "dislikes" noises. Because of this noise association with the automobile, this happy vehicle is referred to by the author as a "spawn of the devil." Could this be a defense mechanism?

Due to its own style, and to its obscure birthplace, the book, fortunately, will have little chance for a popular life. Its contemporaries are those that saw the light of day in the eighteenth century, before the advent of the laboratory. Modern psychology attempts to advance, instead of retard as this book glories in doing, accurate and verifiable explanations of human conduct, or behavior.

S. L. CRAWLEY,
Indiana University.

MEHRON K. THOMSON. *The Springs of Human Action.* New York, D. Appleton and Company, 1927. 501 pp.

An exhaustive literary analysis, of native and acquired "springs of human action" written from the viewpoint of popular psychology. The experimental approach is conspicuous by its absence. A bibliography of some 250 titles is appended.

CARL E. SEASHORE. *Learning and Living in College.* University of Iowa Studies (Series on Aims and Progress of Research), Volume 2, No. 1, 1927. 124 pp.

A simple, straightforward, non-technical presentation of the outcomes of the author's work on the "Gifted Student Project" under the auspices of the National Research Council, during which he visited over 100 American Colleges. The philosophy of the book may be summed up in the dictum:—"Keep each child busy at his highest natural level of successful achievement and he will be happy, useful and good." Modern mass education is not necessarily antagonistic to recognition of the individual. Individual capacity may best be recognized, motivated and rewarded through some fourteen procedures, which, in the author's opinion represent the cream of the procedures noted by him:—a college qualifying examination, departmental placement examinations, sectioning on the basis of ability, project teaching, honors systems, honors time credit for superior work, elimination of competitive introductory courses in favor of basic courses, ablest teacher teaching basic course, various survey and orientation courses, intellectual comradeship of abler students with faculty, character record and motivation, personnel service, junior college, differentiation of functions of institutions—a majority of which appear in the book as chapter heads. No where at the present time will one find so condensed and so readable an account of the aims and philosophy of so many of the newer movements in college education. All who have an interest in higher education might well be urged to read it.

HERBERT A. TOOPS,
Ohio State University.

KIMBALL YOUNG. *Source Book for Social Psychology.* New York, Alfred A. Knopf, 1927. 844 pp.

This addition to the growing list of books on social psychology will serve primarily those readers who lack easy access to the literature since it contains excerpts from 229 sources. The author presents the subject matter of social psychology through the avenue of the group rather than through the individual. In general, the plan followed is, (1) a consideration of pre-human and human behavior as determined by the group; (2) an exposition of the bases of social behavior as revealed through physiological and psychological studies; and (3) an examination of personality as it develops in group relationships upon the triple foundation of social group, culture, and the individual.

To achieve the above plan the author divides the selections under six main headings: general features of social behavior; psychological foundations of social behavior; personality and social behavior; social attitudes and the subjective environment; leadership and prestige in social behavior; and collective behavior. There are 27 chapters, each with an introductory discussion indicating the problems reviewed and explaining the inclusion of certain selections. Questions and exercises and a selected bibliography follow at the end of each chapter. These introductory discussions and class exercises form an integral part of the volume.

While we are informed by the author that the compilation is *selective* and not exhaustive, and despite the praiseworthy attempt to organize the excerpts under six categories of social behavior, one lays down the volume with a growing conviction that social psychology is an amorphous subject; an affair of theory and little experiment. Admitting that the subject as it now exists is not well-defined we may recognize that this source book is truly representative. We find in it, for example, brief gleanings from the writings of anthropologists, neurologists, professors of education, philosophers, physiologists, political scientists, psychologists, sociologists, and a novelist; the greater number of selections being attributed to the sociologists. Not more than 13 per cent of the articles contain data derived from experiments and case studies, or, regarded somewhat differently, more than 85 per cent of the pages are theoretical in nature. Undoubtedly the number of non-experimental papers written under the title of social psychology is far greater than that representing the quantitative or experimental reports, but the omission from this book of many recent important experimental articles detracts from its usefulness, and contributes to its chief weakness which is failure to exploit the newer psychological approach to the field. Nevertheless, Young's "Source Book," useful to the general reader, is a valuable reference work which promises to supplement the content of courses in social psychology.

CHARLES BIRD,
University of Minnesota.

HARRY J. BAKER. *Characteristic Differences in Bright and Dull Children.* Public School Publishing Company, Bloomington, Ill., 1927. 118 pp.

Before intelligence quotients and mental ages can have much bearing upon our school practices they must be translated from figures into facts. These facts must be stated in terms of aptitude for school work, attitude toward study, methods of learning, rate of progress, etc. This has been done for the subnormal child in several recent books, and to some extent has been attempted for the gifted. The dull and the bright, however, have not received their due share of attention. The author has written this little volume in an attempt to interpret the IQ's which lie between 70 and 90 and those between 110 and 130 in terms of school work and procedure. The former group he calls "dull" and the latter "bright," and the book is concerned with contrasting these two groups with respect to their natures and behavior in various school situations.

The material which is the basis for the book was obtained by asking five hundred Detroit teachers to report upon these two groups of children with respect to "differences as they expressed them in psychological terminology, upon differences as expressed in their methods of teaching procedures and in any other way they wished." The author tells us that the reports were remarkably unanimous, but he does not summarize them except as he reflects them in the course of the book. At least one reader is disappointed in getting no better view of the original data upon which the book is based.

The opinions of five hundred teachers ought to furnish an excellent beginning on the problem set for this volume, by providing indications of topics on which research might prove fruitful. Opinions, however, ought not to be considered sufficient basis for a treatise on a subject concerning which facts are obtainable. As a consequence of using opinions instead of facts, the author has often been guilty of dogmatism. For this he occasionally expresses regret, but the fault lies fundamentally with the error of using opinions as facts, and seems indefensible to at least one reader. The author has many opinions which are plausible, but others are highly dubious and controversial. For example, in dealing with the problem of special abilities and disabilities, we are assured that a bright child "seldom suffers from a disability," while a dull child with a special disability has a "serious handicap." On the other hand the combination of a bright child and a special talent "furnishes excellent results," but "a special talent is often a liability to the dull pupil." It is doubtful whether such unverified opinions ought to appear in the field of educational literature, especially when verification is possible.

Part II of the book is devoted to the differences of the two types with respect to their work in specific school subjects. Here also we find a great deal of opinion and a very small quantity of fact. We are told that in the social studies that the dull child exhibits a narrow and selfish point of view, while the bright child is the opposite. In music the dull child waits for others to get the pitch or to lead, while the bright child has all the qualities of a good musician. Dull pupils are "inferior in art;" while "work and study in art have always been a natural gift of bright minds." So throughout the book we find many such statements, and most of them are entirely unsupported by experimental evidence. It may well be true that a large number of the opinions are correct, but it is equally probable that some are doubtful and others misleading.

Part III is a brief discussion of how classes are sectioned in Detroit, and some comments on the types of procedures which the author believes to be necessary for each type of child.

JULIA A. KIRKWOOD. *The Learning Process in Young Children.* University of Iowa Studies. Studies in Child Welfare. Published by the University, Iowa City, 1920. 107 pp.

This monograph gives an account of a study of the learning process in young children. It not only makes a contribution to information on how children learn, but also adds to the methodology of experimentation in this field. The various problems studied center around the task of learning a series of twenty associations which the author has cleverly devised. These consist of associations between blocks and pictures which are not too difficult for children above three years (or even somewhat younger), but which are difficult enough so that only one of seven adults of superior mentality made a perfect score on the first attempt. The twenty associations were to be formed within twenty trials, and the associations were declared to be learned whenever the subject could make three perfect performances in succession. The author fully demonstrates that this criterion is much preferable to the more frequent one of a single perfect performance, or even two perfect performances. Furthermore she has shown that requiring more than three successive, perfect performances would be unnecessary. The number of children tested was 203, but for various reasons such as absence, etc., the number of records used in the report is only 180. The precaution of the author in making the conditions of learning standard and in rejecting all cases which did not conform to these standards is highly commendable.

A number of problems were studied: learning on successive days versus learning on alternate days; plateaus; relearning after an interval of a year; transfer and interference of a set of learned associations with a second series of similar associations; differences between children accustomed to psychological tests and those unaccustomed to such tests; correlations between success in learning these associations and success on various intelligence and form board tests; and other points of interest. In all instances the data have been collected and handled with precision and good judgment. The author very wisely refrains from attempting to establish norms for the various age groups, and from drawing inferences and conclusions beyond those justified by the data. Among the things of particular interest to the reviewer was the contrast between paired children, half of whom were learning on successive days and half on alternate days. In spite of the immaturity of the children the superiority of learning on alternate days was clearly demonstrated for this material. Another point of interest was the fact that plateaus occurred more frequently among the less mature children than among the older ones.

The whole study is interesting and well worth the time of any individual interested in this field of experimental work.

STUART M. STOKE,
Ohio University.

NEW BOOKS AND PAMPHLETS RECEIVED*

Books and pamphlets for review should be sent to James P. Porter, Department of Psychology, Ohio University, Athens, Ohio.

The Classification of Pupils in Private Schools. By The Committee on Intelligence Tests of The Private School Teachers' Association, Philadelphia, Pa. Pamphlet No. 2, 1927. 24 pp.

The Effect of Music. Max Schoen. Harcourt, Brace & Company, New York City. $5.00. 273 pp.

An Elementary Psychology—Revised. D. E. Phillips. Ginn & Company, Boston, Mass. $1.72. 420 pp.

Form Relations Group Test. National Institute of Industrial Psychology, 320 High Holborn, London.

Hearing Through the Sense Organs of Touch and Vibration. Robert H. Gault.

Interpretation of Educational Measurements. Truman Lee Kelley. World Book Company. $2.20. 303 pp.

Kuhlmann-Anderson Intelligence Tests. F. Kuhlmann and Rose G. Anderson. The Educational Test Bureau, University at Fifteenth Avenue, Minneapolis, Minn.

Life, Letters and Labours of Francis Galton. Two volumes. Karl Pearson. The University Press, Cambridge, England.

Morals in Review. A. K. Rogers. Macmillan Company. $3.50. 456 pp.

Practice Exercises in Supervised Study and Assimilative Reading. J. A. Wiley. Iowa State Teachers College, Cedar Falls, Iowa. 112 pp.

Silent Reading and Study Skills. James A. Wiley. Iowa State Teachers College, Cedar Falls, Iowa. 20 cents. 15 pp.

Silent Reading and Study Objectives and Principles. James A. Wiley. Iowa State Teachers College, Cedar Falls, Iowa. $2.00. 308 pp.

The Social Sciences and Their Interpretations. William F. Ogburn and Alexander Goldenweiser. Houghton Mifflin Company. $3.50. 506 pp.

Ten Years of Psychological Research in Schools for the Blind. Samuel P. Hayes. Pennsylvania Institution for the Instruction of the Blind, Philadelphia. 16 pp.

---

* Mention here does not preclude further comment.

INDEX OF AUTHORS

The names of authors of original contributions are printed in CAPITALS AND SMALL CAPITALS.

ADAMS, HENRY F., 25.
Anderson, Amos C., 177, 326.
ARTHUR, GRACE, 251.
BATHURST, J. E., 155.
BEAN, CHARLES HOMER, 369.
BEAR, ROBERT M., 381.
BERGEN, HAROLD B., 42.
Bird, Charles, 533.
BROTEMARKLE, R. A., 415.
Burnham, William H., 242.
Burtt, Harold E., 325.
Crawley, S. L., 243.
CUSHING, HAZEL M., 1.
CUSHMAN, C. L., 500.
DANDISON, BASIL, 483.
Doll, Edgar E., 171.
DOWNEY, JUNE E., 52, 95.
ELWOOD, R. H., 109.
English, Horace B., 246.
Fenton, Jessie C., 174.
FENTON, NORMAN, 394.
FILTER, RAYMOND O., 58.
FOX, EDNA J., 127.
Franken, Richard R., 409.
FRYER, DOUGLAS, 212.
FURFEY, PAUL HANLY, 47.
GARTH, THOMAS R., 268.
GASKILL, PETER C., 394.
GRAUER, DAVID, 297.
Griffitts, C. H., 407.
HOLLINGWORTH, LETA S., 385.
Holman, Portia, 519.
HURLOCK, ELIZABETH B., 490.
HUSBAND, R. W., 465.
KORNHAUSER, ARTHUR W., 85.
KUHLMANN, F., 181.
LASLETT, H. R., 8.
LEHMAN, HARVEY C., 33, 106, 202.
LINCOLN, EDWARD A., 204.
Link, Henry C., 327.
Mateer, Florence, 322.
MILES, W. R., 465.
Morton, R. L., 239.
NEIFELD, M. R., 73.
NIXON, H. K., 135.
Peyton, Edith M., 324.
PORTER, JAMES P., 394.
POULL, LOUISE E., 68.
RENSHAW, SAMUEL, 226.
ROOT, W. T., 297.
Rowles, Emmett, 248.
RUCH, G. M., 1.
SKAGGS, ERNEST B., 503.
SOUTH, EARL B., 348. 437.
STEIN, MARTIN L., 301.
Stoke, Stuart M., 535.
STRONG, EDWARD K., 8, 331.
STUMP, N. F., 117.
TEAGARDEN, FLORENCE M., 280.
Thom, D. A., 247.
Toops, Herbert A., 532.
UHRBROCK, RICHARD S., 95.
WALLIN, J. E. W., 143.
WILLIAMS, HERBERT D., 276.
WITTY, PAUL A., 33, 106, 202.
Worcester, D. A., 240, 329, 411.

INDEX OF SUBJECTS

Titles of books are followed by the name of the author of the work reviewed. All other references are to original articles.

Ability *from interests*, predicting, 212; ——, trial with criteria of the MacQuarrie test of mechanical, 391.

Academic grades, Thorndike intelligence tests and, 207.

Advertisements, experiment on *value of size and repetition*, 483; ——, study of perception of, 135; ——, want versus commodity, 8.

Amount of work one can do, estimates of, 58.

Analyzing athletics, job-, 369.

Aptitude placement test, predictive value of Iowa physics, 381.

Arithmetic mean and the standard deviation, note on some theorems concerning, 73.

Athletics, job-analyzing, 369.

Binet scale, further note on scattering in, 143.

Blood Indians, intelligence of mixed, 208.

Boys from their photographs, judging the intelligence of, 394.

Boys' chums, factors influencing the selection of, 47.

Career: the nurse and college girl, rôle of personality traits in selecting a, 199.

Case for and against psychic belief, the (A symposium), 243.

Character traits in delinquent girls, 1.

Characteristic *Differences* in Bright and Dull Children (Baker), 533.

Child, observations on a visually preoccupied, 52.

Children, diagnostic value of group tests as determined by the *qualitative differences* between normal and feeble-minded, 127; ——, study of self-ratings by, 490; —— with specific reading disability from other non-readers, attempt to sort, 251.

Chums, factors influencing the selection of boys', 47.

Clerical worker and his permanence on the job, social status of, 42.

College-adult level, individual mental testing at, 415; —— freshmen, results from a quantitative questionnaire on likes and dislikes of, 85; —— girl, rôle of personality traits in selecting a career: the nurse and, 199.

Color as variables in two to six kinds, sorting packs of sixty cards with form and, 465.

Committee work, some psychological aspects, 348, 437.

Concepts and methods used in mental testing, critical comments on, 503.

Congress of psychotechnique, fourth international, 519.

Correlation schema, partial, 155.

Criteria of the MacQuarrie test of mechanical ability, trial with, 301.

Delinquency, truancy and, 276.

Delinquent girls, character traits in, 1.

Deviation, some theorems concerning the arithmetic mean and standard, 73.

Differences between normal and feeble-minded children, diagnostic value of group tests as determined by the qualitative, 127.

Disability from other non-readers, attempt to sort children with specific reading 251.

Dislikes used with a group of college freshmen, results from a quantitative questionnaire on likes and, 85.

E——; ten years after the initial report, subsequent history of, 385.

Early childhood education (Pickett and Boren), 170.

Educational psychology (Starch), 520.

Employe representation (Burton), 175.

Environment and the IQ, change of, 289.

Essentials of mental measurement, the (Brown and Thomson), 230.

Executives, vocational guidance of, 331

Experiment in logical learning, class-room, 117.

Experiments on the value of size and repetition in advertisements, 483

Factors influencing the selection of boys' chums, some, 47.

Factors of human psychology (Cole), 328.

Fear: the autobiography of James Edwards (Oliver), 406.

Feeble-minded children, diagnostic value of group tests as determined by the qualitative differences between normal and, 127.

Form board, tentative standards for the Lincoln Hollow Square, 261.

Freshmen, results from a quantitative questionnaire on likes and dislikes of college, 85.

Function of the movies, compensatory, 33; —— of the Sunday funny paper, compensatory, 202.

Funny paper, compensatory function of the Sunday, 202.

Game, clinical values of the Rhode Island intelligence test and the Town picture, 68.

Genetic studies in genius (Terman), 171.

Girl, rôle of personality traits in selecting a career; the nurse and the college, 109.

Girls, character traits in delinquent, 1.

Grades, Thorndike intelligence tests and academic, 297.

Group tests as determined by the qualitative differences between normal and feeble-minded children, diagnostic value, 127.

Growth, periodicity and, 100.

Guidance of executives, vocational, 331.

How to study in college (Headley), 409.

Indians, intelligence of mixed blood, 208; ——, will-temperament of, 512.
Individual mental testing at college-adult level, college student personnel problems, 415.
IQ, change of environment and the, 280.
Intelligence of boys from their photographs, judging, 394; —— test and the Town picture game, clinical values of the Rhode Island, 08; —— tests and academic grades, Thorndike, 207.
Interests, predicting ability from, 212.
International congress of psychotechnique, fourth, 510.
Introduction to social psychology, and (Bernard), 240.
Iowa physics aptitude placement test, predictive value, 381.
Isolation of material and by repetition, memory as affected by, 25.

Job, social status of clerical worker and his permanence, 42.

Learning, classroom experiment in logical, 117; —— of "paired associates," experiment on, 226.
Learning and living in college (Seashore), 532.
Learning how to study and work effectively (Book), 141.
Learning process in young children, the (Kirkwood), 535.
Life of mind, the (Jordan), 531.
Likes and dislikes of a group of college freshmen, results from quantitative questionnaire on 85.
Lincoln hollow square form board, tentative standards for, 264.
Logical learning, class-room experiment in, 117.

MacQuarrie test of mechanical ability, trial with criteria of, 391.
Mean and standard deviation, note on some theorems concerning the arithmetic, 73.
Measurement of advertising effects, the (Hotchkiss and Franken), 325.
Mechanical ability, trial with criteria of the MacQuarrie test, 391.
Mental testing at college-adult level, college student personnel problems: individual, 415; —— testing, critical comments on certain prevailing concepts and methods, 503.
Mental tests, new method of weighting and scaling, 181; —— tests as used in Oklahoma City, study of the reliability of, 509.
Mental Tests in Clinical Practice (Wells), 322.
Mixed blood Indians, intelligence of, 208.
Movies, compensatory function of, 33.

Non-readers, attempt to sort children with specific reading disability from other, 251.
Normal and feeble-minded children, diagnostic value of group tests as determined by the qualitative differences between, 127.

Normal mind, the (Burnham), 170.
Nurse and the college girl, rôle of personality traits in selecting a career, 100.

Oklahoma City, study of the reliability of mental tests used in, 509.

Paired associates, experiment on learning of, 220.
Perception of advertisements, study of, 135
Personality traits in selecting a career; the nurse and college girl, rôle of, 100.
Personnel problems, college student, 415.
Photographs, judging the intelligence of boys from their, 304.
Physics aptitude placement test, predictive value of Iowa, 381.
Picture game, clinical values of the Rhode Island test and the Town, 68.
Picture-story reading lessons (Smith and Courtis), 323.
Placement test, predictive value of Iowa physics aptitude, 381.
Preoccupied child, observations on a visually, 52.
Principles of employment psychology (Burtt), 326.
Problems, college student personnel, 415.
Psychological aspects of committee work, some, 348, 437.
Psychology and education (Ogden), 328.
Psychology for teachers (Benson, Lough, Skinner and West), 410.
Psychology of mental disorders, the (Myerson), 520.
Psychology of selecting men, the (Laird), 324.
Psychology of selling and advertising, the (Strong), 242.
Psychologies of 1925 (Murchison and others), 327.
Psychotechnique, fourth international congress of, 510.

Qualitative differences between normal and feeble-minded children, diagnostic value of group tests as determined by, 127.
Questionnaire on likes and dislikes with group of college freshmen, quantitative, 85.
Quotient, change of environment and the intelligence, 289.

Ratings by children, study of self-, 400.
Reading disability from other nonreaders, attempt to sort children with specific, 251.
Readings in educational psychology (Skinner, Gast and Skinner), 239.
Reliability of mental tests as used in Oklahoma City, study of, 509.
Reluctantly told (Hillyer), 530.
Repetition in advertisements, experiments on the value of size and, 483; ——, memory as affected by isolation of material and by, 25.
Research for teachers (Buckingham), 169.
Rhode Island intelligence test and the Town picture game, clinical values of, 68.

Scale, further note on scattering in Binet, 143.

Scaling mental tests, new method of weighting and, 181.
Scattering in the Binet scale, 143.
Schema, partial correlation, 155.
Scientific advertising (Kitson), 174.
Selection of boys' chums, some factors influencing, 47.
Self-ratings by children, study of, 400.
Selling at retail (Pelz), 407.
Sex differences in the growth of American school children (Lincoln), 520.
Shell shock and its aftermath (Fenton), 240.
Size and repetition in advertisements, further experiments on value, 483.
Sorting packs of sixty cards with form and color as variables in two to six kinds, on, 405.
Source book for social psychology (Young), 532.
Springs of human action, the (Thomson), 531.
Standard deviation, some theorems concerning the arithmetic mean and, 73.
Standards for the Lincoln hollow square form board, 204.
Statistics in psychology and education (Garrett), 237.
Student personnel problems, college, 415.

Temperament test, non-verbal will-, 95.
Temperament and race (Porteus and Babcock), 247.
Temperaments of Indians, will-, 512.
Test and the Town picture game, clinicial values of the Rhode Island intelligence, 68; ——, non-verbal will-temperament, 95; —— of mechanical ability, trial with criteria of the MacQuarrie, 391; ——, predictive value of the Iowa physics aptitude placement, 381.
Testing at the college-adult level, individual mental, 415; ——, critical comments on certain prevailing concepts and methods in mental, 503.
Tests and academic grades, Thorndike intelligence, 297; —— as determined by qualitative differences between normal and feeble-minded children, diagnostic value of group, 127; ——, new method of weighting and scaling mental, 181; —— as used in Oklahoma City, study of the reliability of mental, 509.
Theorems concerning the arithmetic mean and the standard deviation, note on, 73.
Town picture game, clinical values of the Rhode Island intelligence test and, 68.
Traits in delinquent girls, character, 1; —— in selecting a career; the nurse and college girl, the rôle of personality, 190.

Value of the Iowa physics aptitude placement test, predictive, 381.
Values of the Rhode Island intelligence test and the Town picture game, clinical, 68.
Variables in two to six kinds, sorting packs of sixty cards with form and color as, 465.

Visually preoccupied child, observations on, 52.

Want versus commodity advertisements, 8.

Weighting and scaling mental tests, new method of, 181.

Will-temperament test, non-verbal, 95.

Work one can do, estimates of amount, 58; ——, some psychological aspects of committee, 348, 437.

Worker and his permanence on the job, social status of, 42.